

FernUni Hagen
Praktische Informatik
IT-Sicherheit



Masterarbeit zum Thema:

**Echtzeitanalyse von Cyberangriffen in KI-basierten SIEM-Systemen (Detektion) -
Entwicklung eines IT-Grundschutz - Bausteins für KI-Monitoring**

Erstgutachter/in: Univ.-Prof. Dr. Beecks
Zweitgutachter/in: Univ.-Prof. Dr. Keller

vorgelegt von:

Angela Baruth
Max-Bruch-Weg 3
17033 Neubrandenburg
Deutschland

Matrikelnummer: q4470389
Fachsemester: Praktische Informatik - IT-Sicherheit

Datum: 31.12.2024

Inhaltsverzeichnis

Abbildungsverzeichnis in Kurzform

1. Einleitung	1
1.1. Motivation.....	1
1.1. Methodik, Zielsetzung und Abgrenzung.....	1
1.3. Aufbau der Arbeit.....	3
2. Kernkomponente von BSI	4
2.1. Managed Security Information and Event Management.....	5
2.1.1. SIEM mit den wichtigsten Komponenten und Aspekte	5
2.1.2. SIEM Architektur	7
2.1.3. SIEM Integration im Unternehmen.....	9
2.1.3.1. SIEM Team mit organisatorische und technische Vorbereitungen	10
2.1.3.2. SIEM Prozesse	10
2.1.3.3. SIEM Prozesse in Unternehmen durchführen	11
2.1.4. SIEM Logfileanalyse	14
2.1.4.1. Arten von Logfiles	15
2.2. SIEM - Next-Generation.....	16
2.2.1. Herausforderung Globalisierung und technischen Systemen	17
2.3. SIEM Tools	18
2.3.1. Security Monitoring Anwendungen	19
2.3.2. IBM QRadar	19
2.3.2.1. wichtigsten Funktionen von IBM QRadar.....	20
2.3.2.2. IBM QRadar Suite.....	21
2.3.2.3. IBM QRadar Suite mit KI.....	22
2.3.2.4. Demo von IBM QRadar.....	28
2.3.3. Logpoint	28
2.3.3.1. wichtige Funktionen von Logpoint.....	29
2.3.3.2. Logpoint mit KI	33
2.3.3.3. Use Case Logpoint	36
2.3.3.4. Demo von Logpoint.....	38
2.3.4. LogRhythm.....	39
2.3.4.1. wichtige Funktionen von LogRhythm	39
2.3.4.2. LogRhythm mit KI	40
2.3.4.3. Demo von LogRhythm	42
2.3.5. SolarWinds.....	43
2.3.5.1. wichtige Funktionen von SolarWinds mit KI.....	43
2.3.5.3. Demo von SolarWinds	44
2.3.6. ManageEngine.....	45
2.3.6.1. wichtige Funktionen von ManageEngine	45
2.3.6.2. ManageEngine mit KI.....	46
2.3.6.3. Demo von ManageEngine	49
2.3.7. Splunk.....	49
2.3.7.1. wichtige Funktionen von Splunk mit KI	50
2.3.7.2. Splunk mit KI	53
2.3.7.3. Demo von Splunk.....	55
2.3.8. Vergleiche der vorgestellten Tools.....	56
3. KI Anwendungen Monitoring.....	57
3.1. KI Grundlagen.....	57
3.1.1. KI Haupttypen	57
3.1.2. KI Technologien	57
3.1.2.1. Machine Learning (ML).....	58
3.1.2.1.2. ML mit Neural Network	61
3.1.2.1.3. Arten von Neural Network.....	64
3.1.2.1.4. Arten von Algorithmen	74
3.1.3. Deep Learning	85

3.1.4. Natural Language Processing.....	86
3.2. KI- Lebenszyklus mit Bias.....	87
3.3. KI Ethik und Herausforderungen.....	96
3.4. KI Use Case.....	96
3.4.1. KI in Monitoring bei der Cyber-Überwachungen.....	97
3.4.2. DDos.....	99
3.4.3. Ransomware und Viren.....	99
3.4.4. Phishing Angriffe.....	100
3.4.5. Zero-Day-Exploits und Schwachstellenerkennung.....	101
3.4.6. RCDevs als eine Sicherheitsstrategie.....	102
3.5. KI/ML positive und negative Auswirkungen auf Monitoring.....	103
3.5.1. KI/ML positive Auswirkungen auf Monitoring.....	103
3.5.2. KI/ML negative Auswirkungen auf Monitoring.....	104
3.5.3. KI Vergleiche Anwendungsleistungsmanagement vs. IT Operations Analytics.....	105
3.6. KI-Tools.....	106
3.7. KI-Verordnung der EU.....	106
3.8. Risikoanalyse-Bewertung und Vorbeugung von KI.....	112
3.8.1. Risikoanalyse-Erstellung.....	117
3.9. Erstellung von gesetzlichen Anforderungen nach BSI-Schema.....	120
3.10. KI Trends in der Informatik.....	121
4. Rückblick und Schlussbetrachtung.....	123
Anhang I Risikoanalyse KI.....	126
Anhang II KI-IT-Grundschutz-Bausteine nach BSI Schema.....	126
Anhang III Glossar - Begriffserklärungen.....	168
Anhang IV Abkürzungsverzeichnis.....	186
Anhang V Abbildungsverzeichnis.....	190

Abbildungsverzeichnis in Kurzform

Abbildung 1: Kernbestandteile der Cyber.....	4
Abbildung 2: SecurityInformation and Event Management (SIEM).....	5
Abbildung 3: SIEM- Funktionen und Service.....	5
Abbildung 4: SIEMs Architektur.....	7
Abbildung 5: Darstellung einer SIEM-Architektur.....	8
Abbildung 6: Systemkomponenten eines SIEM.....	8
Abbildung 7: Workflow eines SIEM.....	9
Abbildung 8: SIEM Maturity Hierachy.....	9
Abbildung 9: Team SIEM.....	10
Abbildung 10: SIEM-Prozesse.....	10
Abbildung 11: Roadmap Erstellung SIEM Grobkonzept nach CBT Training & Consulting GmbH.....	12
Abbildung 12: Kickoff- Porgrammplan.....	12
Abbildung 13: am häufigsten delegierten SOC-Anwendungsfälle.....	12
Abbildung 14: Projektschritte für die Einführung von SIEMs nach CBT Training & Consulting GmbH.....	13
Abbildung 15: Angriffserkennungen von SIEMs nach CBT Training & Consulting GmbH	14
Abbildung 16: Übersicht SIEM & SOAR.....	17
Abbildung 17: Gartner Bewertung von Tools 2023.....	19
Abbildung 18: IBM QRadar.....	20
Abbildung 19: IBM QRadar Types of content extensions.....	21
Abbildung 20: AQL query flow.....	22
Abbildung 21: Simple AQL queries.....	23
Abbildung 22: QRadar Architekture.....	23
Abbildung 23: Problem insights overview.....	24
Abbildung 24: Metric-based machine learning on z/OS.....	25
Abbildung 25: Log-based machine learning on Linux.....	26
Abbildung 26: IBM Security QRadar SIEM Demo.....	28

Abbildung 27: Logpoint Plattformen.....	28
Abbildung 28: Erkennung kompromittierter Benutzer-Anmeldedaten.....	29
Abbildung 29: Nachverfolgung von Systemänderungen.....	29
Abbildung 30: Erkennung von ungewöhnlichem Verhalten bei privilegierten Konten.....	29
Abbildung 31: Sicherheit für cloudbasierte Anwendungen	30
Abbildung 32: Erkennung von Phishing-Angriffen	30
Abbildung 33: Überwachung von Auslastung und Verfügbarkeit.....	30
Abbildung 34: Logdaten-Management.....	30
Abbildung 35: SIEM für GDPR, HIPAA oder PCI-Compliance	31
Abbildung 36: Suche nach Bedrohungen (Threat Hunting)	31
Abbildung 37: SIEM für die Automatisierung	31
Abbildung 38: Logpiont SIEM	32
Abbildung 39: Logpoint SIEM Platform Solution.....	33
Abbildung 40: wichtige Anwendungsfälle	34
Abbildung 41: SIEM vs. UEBA.....	35
Abbildung 42: UEB- Erkennung	36
Abbildung 43: Best Practices for UEBA.....	36
Abbildung 44: DarkGate Infektionskette	37
Abbildung 45: Logpoint AgentX Isolate-Unisolate Host	38
Abbildung 46: Logpiont SIEM Demo.....	38
Abbildung 47: LogRhythm.....	39
Abbildung 48: Evolution of SIEM-Software.....	40
Abbildung 49: LogRhythm Architektur	42
Abbildung 50: LogRhythm SIEM Demo	42
Abbildung 51: SolarWinds SIEM.....	43
Abbildung 52: künstliche Intelligenz für IT-Abläufe bei SolarWinds.....	44
Abbildung 53: SolarWinds SIEM Demo	44
Abbildung 54: ManageEngine.....	45
Abbildung 55: Analyse des Benutzer- und Entitätsverhaltens mithilfe von KI und Prozessflussdiagramm für die Analyse von Benutzerentitäten und -verhalten.....	48
Abbildung 56: Prozessflussdiagramm für die Vorhersage von Ausfällen	48
Abbildung 57: MangeLogs - Audit - Secure – Be Compliant	49
Abbildung 58: ManageEngine Demo	49
Abbildung 59: Splunk Observability Cloud Schema	50
Abbildung 60: Splunk Observability	51
Abbildung 61: Quellen von Spunk	51
Abbildung 62: Risikobewertung bei MLTK- oder Out-of-the-Box-Anwendungsfall	54
Abbildung 63: ML-basierten Analysen in Splunk	54
Abbildung 64: Vorhersage von Datenausfällen in Splunk.....	55
Abbildung 65: Splunk Demo	55
Abbildung 66: drei wichtigsten Arten von KI	57
Abbildung 67: KI-Technologien.....	57
Abbildung 68: Leistungsbestandteile der Künstlichen Intelligenz	58
Abbildung 69: Arten von Machine Learning Algorithmen	58
Abbildung 70: Unsupervised Learning (Unüberwachtes Lernen) ist eine Art von Machine Learning, als eigenständiges Muster und Zusammenhänge in den Daten findet.....	59
Abbildung 71: Überwachtes maschinelles Lernen trainiert Muster und Zusammenhänge anhand von Daten.....	59
Abbildung 72: Semi-überwachten Lernen.....	60
Abbildung 73: einfaches Beispiel von verstärkendem Lernen durch Belohnungen.....	60
Abbildung 74: Maschinelles Lernen im Überblick: Anwendungsbeispiele nach Arten.....	61
Abbildung 75: neuronales Netzwerk	62
Abbildung 76: Einordnung neuronale Netz-Arten	64
Abbildung 77: einfache und Multilayer neuronale Perceptron	65
Abbildung 78: Netzwerkdiagramm eines Feedforward-Netzes.....	65
Abbildung 79: Faltung in Convolutional Neural Networks.....	66

Abbildung 80: Aufbau eines Recurrent Neural Networks und Long Short-Term Memory Units	67
Abbildung 81: Modulare neuronale Netzwerke (MNNs)	67
Abbildung 82: Radialen Basisfunktionen-Neuronale Netzwerke	68
Abbildung 83: Liquid State Machine-Neuronale Netzwerke	69
Abbildung 84: Residuale-Neuronale Netzwerke	69
Abbildung 85: Generative Adversarial Networks	70
Abbildung 86: Self Organizing Maps.....	71
Abbildung 87: Deep Belief Networks	72
Abbildung 88: Restricted Boltzmann Machines	72
Abbildung 89: Autoencoders.....	73
Abbildung 90: Machine Learning nutzt Daten, um Muster und Zusammenhänge in Daten zu identifizieren.....	74
Abbildung 91: Lineare Regression-Algorithmus	75
Abbildung 92: Logistische Regression-Algorithmus	75
Abbildung 93:Vergleich lineare Regression vs. logistische Regression	76
Abbildung 94: Naïve Bayes-Naive Bayes-Klassifikatoren-Algorithmus	76
Abbildung 95: - Support Vector Machine-Algorithmus (SVM) Algorithmus.....	77
Abbildung 96: Entscheidungsstruktur-Algorithmen	77
Abbildung 97: KNN-Diagramm.....	77
Abbildung 98: Clustering-Algorithmus.....	78
Abbildung 99: k-Means Clustering Prozess.....	79
Abbildung 100: Ergebnisse unseres Clusteranalyse-Beispiels. Clusterbildung mit dem DBSCAN-Algorithmus. Auswertung der gefundenen Cluster mit dem Calinski-Harabasz-Index und der Silhouttenmethode	80
Abbildung 101: Clusterbildung mit dem HDBSCAN-Algorithmus. Auswertung der gefundenen Cluster mit dem Calinski-Harabasz-Index und der Silhouttenmethode	80
Abbildung 102: beispielhafte Darstellung eines hierarchischen Clusterings beim Machine Learning	81
Abbildung 103: Random-Forest-Algorithmus.....	82
Abbildung 104: AdaBoost	82
Abbildung 105: Gradient Boosting-Algorithmus.....	83
Abbildung 106: LightGBM	83
Abbildung 107: CatBoost	84
Abbildung 108: XGBoost.....	84
Abbildung 109: Deep neutral network.....	85
Abbildung 110: Machine Learning vs. Deep Learning: der Unterschied liegt in der Feature Extraktion und dem Einsatz von tiefen, künstlichen neuronalen Netzen	86
Abbildung 111: Natural Language Processing	87
Abbildung 112: Unterschiede von NLP, NLU und NLG	87
Abbildung 113: Datenerhebung von Bais	90
Abbildung 114: Entwicklung, Implementierung und Nutzung von Bais.....	90
Abbildung 115: Formaler Aufbau einer KI-Anwendung	91
Abbildung 116: Lebenszyklus einer KI-Anwendung	93
Abbildung 117: Abstrahierter Lebenszyklus einer KI-Anwendung.....	93
Abbildung 118: Training des ML-Modells einer KI-Anwendung.....	95
Abbildung 119: IOT & Maschinelles Lernen in 3 Stufen ML-Ansatz	98
Abbildung 120: wesentliche Unterscheidungsmerkmale zwischen SIEM und SOAR	98
Abbildung 121: Rapid Connectivity Installation and Asset Onboarding.....	98
Abbildung 122: Responsible AI Principles von infotech.....	104
Abbildung 123: Laufe des AI-Acts von 2024 bis 2026	106
Abbildung 124: KI-Regulierungen und ihre Umsetzungen.....	107
Abbildung 125: Hochrisiko-KI-Systeme nach Annex I und III	108
Abbildung 126: KI-Risiko Pyramiede	108
Abbildung 127: Kategorien von KI-Systemen nach dem EU AI-Act	109
Abbildung 128: gesetzliche KI-Verordnung der EU	111
Abbildung 129: internationale Landschaft der KI-Initiativen.....	112

Abbildung 130: Risk Mitigation Essential	112
Abbildung 131: Lebenszyklus des Analysemodells	112
Abbildung 132: Risikoentwicklung	113
Abbildung 133: Risikoklassifizierung.....	114
Abbildung 134: Risk Management Framework.....	114
Abbildung 135: Risk Management Framework.....	115
Abbildung 136: 2 Level Algorithmic Risk Monitor und Management	115
Abbildung 137: ML/AI basiert extrinsisches Risikomanagement	116
Abbildung 138: Algorithmische Risikobewertung und Auswirkungsabschätzung.....	116
Abbildung 139: Risk Governance	116
Figure 140: Künstliche Entscheidungsfreiheit als Grundlage für risikobasierte Governance	117
Abbildung 141: KI-Analyse.....	118
Abbildung 142: AI-Act-Analyse	119
Abbildung 143: KI-Eigenschaften, die auf Richtliniendokumente abgebildet werden.....	120
Abbildung 144: Beziehung zwischen KI-Bedrohungen und Sicherheitskontrollen	120
Abbildung 145: Vergleich klassische Computer vs. Quantenrechner	121
Abbildung 146: Übersicht von der Cyber-Security	168
Abbildung 147: NKCS-Verbund	169
Abbildung 148: Übersicht über BSI-Publikationen zum Sicherheitsmanagement	171
Abbildung 149: Aufbau der Energiesynchronisationsplattform zur Automatisierung und Standardisierung des Energieflexibilitätshandels	173
Abbildung 150: Key Concepts of COBIT 5	174
Abbildung 151: ITIL-Zertifizierung von Service-Management-Systemen	175
Abbildung 152: NIST Cyber Security Framework 2.0	175
Abbildung 153: DISA-Überblick	176
Abbildung 154: CIS	177
Abbildung 155: ACSC-Logo.....	177
Abbildung 156: BSI-Schutzziele ISO/IEC 27001	178
Abbildung 157: Gefährdungen - Schutzziele - Schutzbedarfe.....	178
Abbildung 158: Risikomatrix mit Risikoeinstufung	179
Abbildung 159: Zusammenhang zwischen Angriffen auf die Schutzziele und Gegenmaßnahmen.....	180
Abbildung 160: Elemente einer Cyber-Security-Strategie	181
Abbildung 161: Top 10 Cyberangriffe 2023	182
Abbildung 162: Die Lage der IT-Sicherheit in Deutschland 2023 im Überblick und Bedrohungsziele laut BSI.....	182
Abbildung 163: Threat – Asset – Vulnerability – Risk - Zusammenhänge.....	185

1. Einleitung

1.1. Motivation

In der heutigen digitalen Ära stellen Cyberangriffe eine stetig wachsende Bedrohung für Unternehmen und Organisationen dar. Die Fähigkeit, solche Angriffe zu erkennen und darauf zu reagieren, ist entscheidend, um potenziellen Schaden abzuwehren. Security Information and Event Management (SIEM)-Systeme spielen eine Schlüsselrolle in der Cybersecurity-Landschaft, indem sie Daten aus verschiedenen Quellen sammeln, analysieren und in Echtzeit Alarmer generieren. Mit dem Aufkommen fortschrittlicher und komplexer Angriffsmuster stoßen herkömmliche SIEM-Systeme jedoch an ihre Grenzen. Die Integration von Künstlicher Intelligenz (KI) in SIEM-Systeme verspricht, die Effektivität dieser Tools signifikant zu erhöhen. Vor diesem Hintergrund ist das Ziel dieser Arbeit, die Entwicklung und Bewertung eines KI-basierten SIEM-Systems zur Detektion und Analyse von Cyberangriffen in Echtzeit zu untersuchen.

Die drei Hauptelemente des Bundesamts für Sicherheit in der Informationstechnik (BSI) - Prevention, Detection und Response - durchlaufen Prozesse, die darauf abzielen, Cyberangriffe vorzubeugen, herauszufiltern und zu verhindern. Insbesondere wird das Thema Detection wissenschaftlich untersucht, um effiziente organisatorische und technische Strategien zu entwickeln und den Einsatz von KI in Hard- und Software zu optimieren. Dabei liegt der Fokus auf der präventiven Bekämpfung, Erkennung und Wiederherstellung von Cyberangriffen, wobei verschiedene Szenarien und Anwendungsmöglichkeiten detailliert betrachtet werden. Das übergeordnete Ziel besteht darin, eine umfassende Perspektive auf die Erkennung und Vermeidung von Cyberangriffen durch SIEM-Monitoring mit KI zu bieten. Besonders wichtig ist es, die Bedeutung einer neuen KI-Verordnung der Europäischen Union zu betonen, die ab den 01. August 2024 in Kraft tritt und bis Ende 2026 in sämtlichen relevanten Bereichen umgesetzt werden muss. Diese Verordnung sieht eine verstärkte gesetzliche Regulierung von KI-Technologien in verschiedenen Anwendungsbereichen vor, welche in einer Risikobewertung genauer analysiert wird (siehe Vorlage SOC-CMM for CERT).¹ In diesem Kontext werden einige Anforderungen entwickelt, die eine klare Definition und effektive Umsetzung von Standards und Vorschriften für den Einsatz von KI in der Cybersicherheit ermöglichen (in einer Vorlage von SOC-CMM (CERT)).²

Folgende Fragestellungen werden in dieser Masterarbeit untersucht:

- Welche Komponenten sind integraler Bestandteil einer Cybersecurity-Strategie?
- Wie trägt die Detection (Erkennung) dazu bei, die allgemeine Cybersicherheit zu stärken?
- Welche Funktionsweise und spezifischen Tools werden in SIEM verwendet, und wie gestaltet sich ihre Arbeitsweise?
- Inwiefern beeinflusst Künstliche Intelligenz (KI) bei SIEM-Tools die Landschaft der Cybersicherheit?
- Welche Risiken können aus der KI-Verordnung der Europäischen Union von 2024 abgeleitet werden?
- Welche IT-Grundschutz-Bausteine für KI, speziell auch für Monitoring können erstellt werden?

1.1. Methodik, Zielsetzung und Abgrenzung

Die vorliegende Masterarbeit bezieht sich auf eine umfassende Untersuchung der Erkennungstechnologien mit einem speziellen Fokus auf die Anwendung von Methoden aus dem Bereich Security Information and Event Management (SIEM)-Tools. Die methodische Grundlage dieser Analyse beruht auf einer gründlichen Recherche von Informationen und Daten aus verschiedenen Quellen, darunter wissenschaftliche Literatur, Onlineartikel, Tutorien, aktuelle Studien, Seminare, Unternehmensplattformen sowie Whitepapers, die in dieser Masterarbeit einfließen und zur Erstellung einer Risikoanalyse und Entwicklung von IT-

vgl.¹ ff. (SOC-CMM, 2024)

vgl.² ff. (SOC-CMM, 2024)

Grundschutz-Bausteinen, insbesondere für das KI-Monitoring im Rahmen der AI-Act EU, herangezogen werden. Hierbei werden auf einen fundierten Wissensschatz zurückgegriffen, der durch Schulungen bei der Syss GmbH seit 2011 aufgebaut wurde sowie einer SIEMs Schulung für Abwehr von Cyberangriffen hilfreich eingearbeitet werden und nicht nur die Vertiefung des theoretischen Wissens, sondern auch dessen praktische Anwendung durch umfangreiche Tests in verschiedenen Bereichen der IT-Sicherheit. Die Synergie aus theoretischem Verständnis und praktischer Erfahrung fließt in die Ausführungen dieser Arbeit ein, um eine praxisorientierte und fundierte Perspektive zu gewährleisten. Der Anhang dieser Arbeit umfasst eine umfassende Risikoanalyse AI-Act EU sowie die Erstellung von eventuellen IT-Grundschutz-Bausteinen für das KI-Monitoring, welche einen zusätzlichen Mehrwert sowie durch wissenschaftliche Methoden als auch für die Forschungsfrage dieser Masterarbeit bieten. Diese Dokumente erleichtern die Anwendbarkeit der erzielten Erkenntnisse in konkreten Szenarien und stellen eine praktische Ergänzung zum theoretischen Rahmen dar. Das Ziel dieser Masterarbeit besteht darin, einen ganzheitlichen Einblick in ausgewählte IT-Sicherheitskomponenten zu bieten und dabei sowohl theoretische Konzepte als auch praxisorientierte Anwendungen zu beleuchten. Die Abgrenzung fokussiert sich auf den Einsatz von SIEM-Methoden und -Tools in Kombination mit Künstlicher Intelligenz (KI). Dabei wird eine umfassende Risikoanalyse durchgeführt sowie IT-Grundschutz-Bausteine für das Monitoring von KI im gesamten KI-Lebenszyklus entwickelt. Dadurch leistet die Arbeit einen wertvollen Beitrag zum besseren Verständnis und zur praxisnahen Anwendung moderner IT-Sicherheitspraktiken.

1. Detaillierte Betrachtung und tiefgreifendes Verständnis der zentralen BSI-Komponente "Detection" auf Basis von SEIMs-Monitoring und KI-Einsatz, inklusive einer eingehenden Analyse von Merkmalen und Einsatzbereichen.
2. Untersuchung der Integration von Künstlicher Intelligenz im Bereich der Cybersicherheit mit Schwerpunkt auf präventive, erkennende und reaktive Sicherheitsmaßnahmen in SIEM-Tools.
3. Eine abgespeckte Risikoanalyse (im Anhang I) und daraus ergebende Entwicklung von IT-Grundschutz-Bausteinen für das KI-Monitoring nach der EU-KI-Verordnung (im Anhang II).
4. Erklärung der Begrifflichkeiten der IT-Sicherheit (im Anhang III Glossar).

theoretische Grundlagen und Forschungskonzeptfragestellungen:

Fragen zu Detection

1. Was ist ein SIEM und welche Komponenten sind für seine Funktionsweise entscheidend?
2. Welche IT-Architekturen eines SIEM-Systems gibt es und wie diese in Unternehmen integriert werden?
3. Wie funktioniert die Logfileanalyse in einem SIEM-System und welche Informationen können daraus gewonnen werden?
4. Welche sind die wichtigsten SIEM-Tools auf dem Markt und wie unterscheiden sich diese voneinander?

Fragen zu KI-Anwendungen Monitoring

1. Was sind die Grundlagen von Künstlicher Intelligenz und wie werden diese im Bereich der IT-Sicherheit angewendet?
2. Welche Use Cases für Künstliche Intelligenz im Monitoring von IT-Systemen und Netzwerken können vorhanden sein?
3. Welche Inhalte der Verordnungen hat die EU bezüglich Künstlicher Intelligenz erlassen und wie beeinflussen diese die Entwicklung und Anwendung von KI?
4. Welche Trends gibt es in der Nutzung von Künstlicher Intelligenz im Bereich der Informatik und wie wirken diese sich auf das Monitoring von IT-Systemen aus?

Fragen zu Erstellung von KI IT-Grundschutz-Bausteine SIEMs Monitoring

1. Wie geht man bei der Risikoanalyse im Zusammenhang mit Künstlicher Intelligenz vor und welche spezifischen Risiken können auftreten?
2. Welche IT-Grundschutz-Bausteinen im Kontext von Künstlicher Intelligenz können von der Risikoanalyse abgeleitet werden, hinsichtlich des SIEMs-Monitorings?

Fragen zu Begriffserklärungen

1. Was sind nationale und internationale Standards in der IT-Sicherheit und warum sind diese wichtig?
2. Welche gesetzlichen Grundlagen bezüglich der IT-Sicherheit nennt das BSI und welche Bedeutung haben sie für Unternehmen?
3. Welche Schutzziele gibt es in der Informationssicherheit und wie tragen diese zum Gesamtschutz bei?
4. Wie definiert das BSI Schwachstellen und welche Arten von Schwachstellen werden unterschieden?
5. Welche Elemente sollte eine umfassende Cyber-Security-Strategie enthalten und warum sind diese relevant?
5. Welche Typen von Angreifern werden im Bereich der IT-Sicherheit unterschieden und wie unterscheiden diese sich in ihren Vorgehensweisen?

1.3. Aufbau der Arbeit

Die vorliegende Masterarbeit zeichnet sich durch eine systematische Untersuchung zentraler Aspekte der IT-Sicherheit aus. Zu Beginn werden die drei Hauptkomponenten des BSI – Prävention, Detektion (mit besonderem Schwerpunkt) und Response – beleuchtet, um ein fundiertes Verständnis ihrer jeweiligen Rollen und Funktionen zu erlangen.

Im Anschluss daran widmet sich die Arbeit einer ausführlichen Analyse der Funktionsweise von Security Information and Event Management (SIEM)-Systemen. Diese Analyse ermöglicht einen tiefen Einblick in die operativen Mechanismen und die Bedeutung dieser essenziellen Sicherheitslösungen. Darauf aufbauend erfolgt eine detaillierte Betrachtung und der Vergleich der am häufigsten genutzten SIEM-Tools. Hierbei werden deren charakteristische Merkmale sowie die unterschiedlichen Einsatzbereiche kritisch untersucht.

Im abschließenden Teil der Arbeit liegt der Fokus auf dem Einsatz Künstlicher Intelligenz (KI) im Bereich des Cybersicherheits-Monitorings. Dabei werden die Anwendungen und potenziellen Auswirkungen dieser innovativen Technologie auf präventive, detektierende und reaktive Sicherheitsmaßnahmen eingehend betrachtet. Ein zentrales Thema ist dabei die Integration neuer Technologien wie der künstlichen Intelligenz in bestehende Sicherheitsarchitekturen.

Zusätzlich wird im Rahmen der Arbeit eine Risikoanalyse mit einer SOC-Analyse gemäß dem AI-Act EU von 2024 durchgeführt (im Anhang I). Basierend darauf werden eigene Anforderungen an KI-Schutzziele sowie der KI-Lebenszyklus, strukturiert nach dem BSI-Schema, entwickelt (im Anhang II).

Abschliessend wird im Anhang III – Glossar relevante Begrifflichkeiten zu dieser Masterarbeit kurz erklärt.

Die Arbeit bietet einen umfassenden Überblick über die wesentlichen Aspekte der KI- IT-Sicherheit, die natürlich noch tiefer betrachtet werden können, um wirklich alle IT-Sicherheitsrisiken abzudecken.

2. Kernkomponente von BSI

Das BSI definiert drei wesentliche Elemente der Cyber-Sicherheit: **Prävention, Erkennung und Reaktion**. Eine Zusammenstellung der Managementtools, die zur Implementierung einer Zero-Trust-Sicherheitsstrategie in einem Unternehmen verwendet werden können wird in der nachfolgende Übersicht kurz erklärt:³



Abbildung 1: Kernbestandteile der Cyber

Prävention ist effektiver als Strafverfolgung. Während der Präventionsphase sollten Sicherheitsrichtlinien, Schulungsprogramme und Zugriffskontrollen entwickelt werden. Die Sicherheitsrichtlinie definiert, was geschützt werden muss, festigt Verantwortlichkeiten und bildet den Grundstein. Sicherheitssensibilisierung klärt Mitarbeiter über Sicherheitsaspekte und ihre Verantwortlichkeiten auf. Zugriff sollte auf "Need-to-know"-Basis gewährt werden, und Zugriffskontrollen sollten Identifikation, Authentifizierung und Autorisierung umfassen. Die Identifikation ist einzigartig, Authentifizierung überprüft sie, und Autorisierung gewährt Zugriff basierend auf festgelegten Regeln. Biometrie ist eine starke Authentifizierungsoption. Autorisierung sollte nach dem Prinzip des minimalen Privilegs erfolgen. Nach Einführung von Richtlinien, Sensibilisierung und Zugriffskontrollen sind Erkennungsstrategien und Reaktionspläne zu implementieren. Eine proaktive Vorbereitung ist besser als eine reaktive Reaktion auf unterschätzte Bedrohungen. Die Erkennung von Missbrauch erfordert mehr als nur Alarme, und die Reaktion auf Vorfälle erfordert effektive Koordination von Ressourcen und Zeit ist dabei entscheidend.⁴

Detection beinhaltet die Erkennung von Systemkompromissen und ist entscheidend, da jedes System trotz Schutzmaßnahmen bei ausreichender Motivation und Fähigkeiten kompromittiert werden kann. Es gibt keine vollständige Sicherheitslösung, daher ist eine Verteidigungsstrategie in Schichten wichtig. **Intrusion Detection Systems (IDS)** überwachen und benachrichtigen bei verdächtigen Aktivitäten, erkennen Angriffssignaturen und Änderungen. Das gesamte System sollte überwacht werden, und IDS-Tools sollten strategisch auf Netzwerk- und Anwendungsebene platziert werden. Die Konfiguration des IDS erfordert Abstimmung auf spezifische Netzwerk- oder Hostmerkmale, um normale von böswärtiger Aktivität zu unterscheiden. Ein gut konfiguriertes IDS ist wie ein Alarm mit Verstand, welcher durch einen dokumentierten Reaktionsplan erstellt werden sollte.⁵

Incident Response ist entscheidend für Sicherheitsvorfälle werden sollte. Ein vorher gut geplanter **Computer Security Incident Response Plan (CSIRP)** mit klaren Rollen und

vgl.³ (Zillmann, et al.)

vgl.⁴ (James LaPiedra, 2002)

vgl.⁵ (Institute, 2000)

Verantwortlichkeiten sind unerlässlich. Die Wahl der Reaktionsmethode, sei es das Abschneiden der Verbindung und Wiederherstellen des Systems oder die Verfolgung des Angreifers, sollte dokumentiert sein. Die Beziehungen zu externen Akteuren wie Strafverfolgung und Medien sollten vor einem Vorfall aufgebaut werden. Nach der Meldung eines Vorfalls müssen Schäden bewertet, das System gereinigt und wiederhergestellt werden. Die Analyse nach dem Vorfall ist entscheidend, um aus Erfahrungen zu lernen und den Informationssicherheitszyklus zu stärken. Ein kontinuierlicher Verbesserungsprozess erfordert disziplinierte Verwaltung und Organisation weitreichende Unterstützung. **"Nicht-wissen-wollen"-Management** kann kostspielig sein, daher ist das Engagement der gesamten Organisation wichtig.⁶

Im Folgenden wird ausschließlich das Thema Detection mit SIEMs im Kontext von Künstlicher Intelligenz in der Technologie betrachtet.

2.1. Managed Security Information and Event Management

Security Information and Event Management (SIEM) ist ein umfassender Ansatz zur Verwaltung der Informationssicherheit einer Organisation. Dabei erfolgt die Sammlung, Analyse und Interpretation von sicherheitsrelevanten Daten aus verschiedenen Quellen. Das Hauptziel von SIEM besteht darin, einen zentralisierten und Echtzeitüberblick über die Informationssicherheit einer Organisation bereitzustellen.



Abbildung 2: Security Information and Event Management (SIEM)

- **SIM** – Protokollsammlung, Archivierung, historische Berichterstattung, Forensik;
- **SEM** – Echtzeit-Berichterstellung, Protokollerfassung, Normalisierung, Korrelation, Aggregation;
- **SIEM** – Protokollsammlung, Normalisierung, Korrelation, Aggregation, Berichterstellung.⁷

2.1.1. SIEM mit den wichtigsten Komponenten und Aspekten

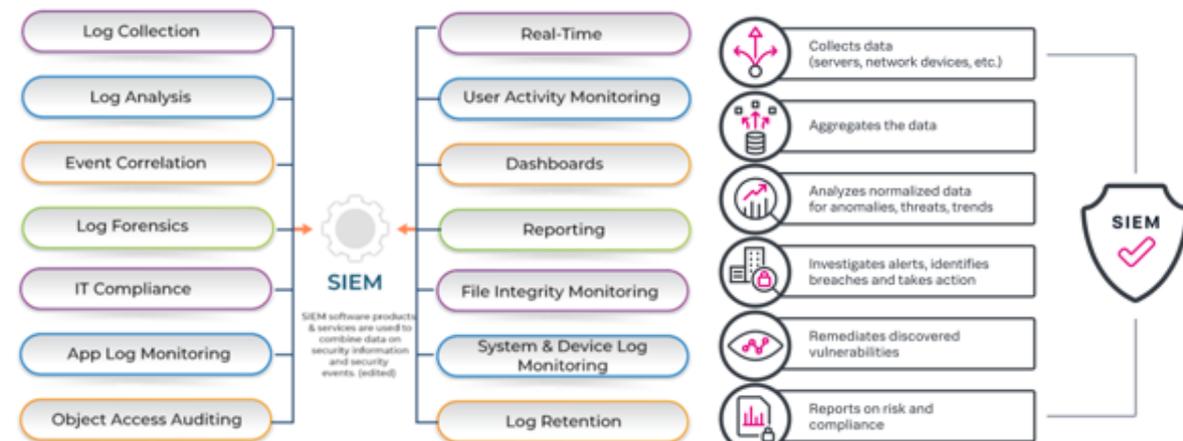


Abbildung 3: SIEM- Funktionen und Service

vgl.⁶ (LaPiedra)

vgl.⁷ (Natalia G. Miloslavskaya, 2018)

Datensammlung ist der erste Schritt in einem umfassenden Sicherheitsinformationen- und Ereignismanagement SIEM-Prozess. SIEM-Systeme spielen eine zentrale Rolle bei der Sammlung und Aggregation von Protokolldaten aus der gesamten technologischen Infrastruktur einer Organisation. Diese Daten reichen von Host-Systemen und Anwendungen bis hin zu Netzwerk- und Sicherheitsgeräten. Sicherheitsereignisse und Vorfälle werden identifiziert, protokolliert und normalisiert, um eine effektive Korrelation und Analyse zu ermöglichen. Danach folgt die **Normalisierung und Korrelation**. SIEM-Tools übernehmen die Aufgabe, Daten aus verschiedenen Quellen zu normalisieren und zu korrelieren. Dies ermöglicht die Identifizierung von Mustern, Trends und potenziellen Sicherheitsbedrohungen. Die Normalisierung umfasst die Umwandlung von Rohdaten in ein konsistentes Format, das für die Analyse geeignet ist. Die **Alarmierung und Incident Response** sind entscheidende Schritte, um auf Sicherheitsbedrohungen angemessen zu reagieren. SIEM-Systeme generieren in Echtzeit Alarme basierend auf vordefinierten Regeln und der Korrelation von Ereignissen. Incident-Response-Funktionen ermöglichen es Sicherheitsteams, schnell auf identifizierte Bedrohungen zu reagieren. SIEM bietet außerdem **Dashboards und Berichtsfunktionen**, die Sicherheitsanalysten und Stakeholdern eine visuelle Darstellung des Sicherheitsstatus bieten. Diese Berichte können angepasst werden, um Compliance-Anforderungen zu erfüllen und die Wirksamkeit von Sicherheitsmaßnahmen zu zeigen. **Compliance-Management** ist ein weiterer Aspekt von SIEM. Lösungen in diesem Bereich unterstützen Organisationen bei der Einhaltung branchenspezifischer Vorschriften und Standards, indem sie Berichte und Dokumentationen bereitstellen. Einige SIEM-Systeme integrieren **User and Entity Behavior Analytics (UEBA)**, um Benutzerverhaltensmuster zu analysieren und Anomalien zu erkennen. Die Integration mit anderen Sicherheitstools wie Antivirensoftware, Firewalls und Intrusion Detection/Prevention-Systemen sind ebenfalls möglich. Mit der zunehmenden Nutzung von **Cloud-Services** bieten moderne SIEM-Lösungen Unterstützung für die Überwachung und Analyse von Protokollen und Ereignissen aus cloudbasierten Umgebungen. Die Integration von **Threat Intelligence** verbessert die Fähigkeit zur Identifizierung und Reaktion auf bekannte Bedrohungen. **Sicherheitsautomatisierung und -orchestrierung** sind Funktionen einiger SIEM-Lösungen, die darauf abzielen, Reaktionsmaßnahmen zu optimieren und zu automatisieren.⁸

Datenquellen (Source Device) umfassen sämtliche Geräte und Systeme in einer IT-Umgebung, die Sicherheitsereignisse generieren können. Das reicht von Firewalls über IDS/IPS-Systeme, Antivirensoftware, Netzwerkgeräte bis hin zu Servern, Anwendungen und Endpunkten. Jedes dieser Geräte spielt eine Rolle bei der Sicherheitsüberwachung, indem es Protokolle und Ereignisse generiert, die später von SIEM analysiert werden. Die **agentenbasierte, agentenlose, API-basierte Logsammlung (Log Collection)** beinhaltet den Prozess der Extraktion von Protokolldaten von den Datenquellen. Hierbei verwendet SIEM verschiedene Ansätze, einschließlich agentenbasierter Methoden, bei denen spezielle Softwareagenten auf den Quellgeräten installiert sind, agentenloser Methoden, bei denen Protokolle direkt von den Geräten gesammelt werden, und API-basierter Methoden, die Schnittstellen für die Datenextraktion nutzen. **Normalisierung (Parsing Normalization)** standardisiert die gesammelten Protokolldaten, um sicherzustellen, dass sie in einer einheitlichen Form vorliegen. Dies ist entscheidend, da verschiedene Geräte unterschiedliche Protokollformate verwenden können. Durch die Normalisierung werden die Daten so angepasst, dass sie effizient von der Regel- und Korrelations-Engine verarbeitet werden können. **Regeln und Korrelation (Rule Engine & Correlation Engine)** definiert vordefinierte Regeln zur Identifizierung verdächtiger Aktivitäten oder Sicherheitsverletzungen. Diese Regeln basieren auf Sicherheitsrichtlinien und Best Practices. Die Korrelations-Engine analysiert Ereignisse in Bezug auf ihre Zusammenhänge und identifiziert komplexe Angriffsmuster. Die Kombination von Regeln und Korrelation ermöglicht eine effektive Erkennung von Sicherheitsvorfällen. **Logspeicher (Log Storage)** umfasst die sichere und effiziente Speicherung der gesammelten und normalisierten Protokolldaten. Diese Protokolle dienen nicht nur der Nachverfolgung von Sicherheitsvorfällen, sondern erfüllen auch Compliance-Anforderungen. Der Log-Speicher kann lokal oder in der Cloud bereitgestellt werden und ermöglicht Sicherheitsteams historische Analysen sowie forensische Untersuchungen bei

vgl.⁸ (WOLF, 2023)

Bedarf. **Überwachung** (Monitoring) ist ein kontinuierlicher Prozess, bei dem das SIEM-System in Echtzeit nach sicherheitsrelevanten Ereignissen sucht. Sicherheitsanalysten überwachen die Aktivitäten, Alarme und Warnungen, die vom SIEM generiert werden. Diese Echtzeitüberwachung ermöglicht es, potenzielle Sicherheitsbedrohungen sofort zu erkennen und darauf zu reagieren, um Angriffe zu stoppen oder Sicherheitslücken zu schließen. Monitoring ist entscheidend für eine proaktive Sicherheitsverteidigung.⁹

Trotz der zahlreichen Vorteile gibt es auch **Herausforderungen** bei der Implementierung von SIEM. Die Prozesse können komplex und ressourcenintensiv sein. Die ordnungsgemäße Anpassung von Regeln und Schwellenwerten ist erforderlich, um falsche Positivmeldungen zu reduzieren. Eine kontinuierliche Überwachung und Aktualisierung von Regeln ist entscheidend, um sich entwickelnden Bedrohungen gerecht zu werden. Die **Entwicklung von SIEM** schreitet voran, indem fortschrittliche Technologien wie maschinelles Lernen und künstliche Intelligenz integriert werden. Dies trägt dazu bei, die Erkennung und Reaktion auf Bedrohungen weiter zu verbessern.

Insgesamt ist die **Implementierung einer SIEM-Lösung** ein entscheidender Aspekt einer umfassenden Cybersecurity-Strategie. Sie unterstützt Organisationen dabei bei der einer umfassende und proaktive Sicherheitsüberwachung, um Sicherheitsvorfälle zu identifizieren, darauf zu reagieren und gleichzeitig eine umfassende Übersicht über die Sicherheitslage zu bieten und um die allgemeine Informationssicherheit zu stärken.¹⁰

2.1.2. SIEM Architektur

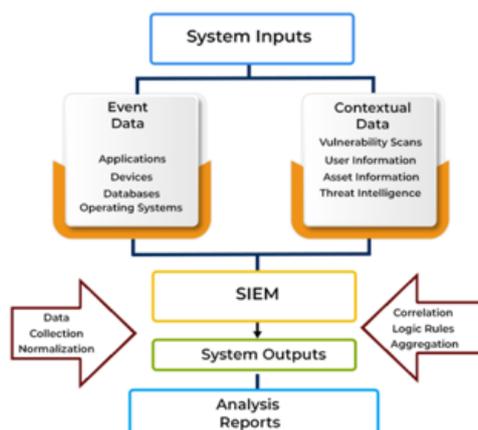


Abbildung 4: SIEMs Architektur

SIEM bildet eine robuste Architektur für die umfassende **Verwaltung von Protokollen und Sicherheitsereignisse**. Diese Lösung spielt eine entscheidende Rolle bei der intelligenten Datensammlung, um Unternehmen aussagekräftige Informationen über Mitarbeiterleistung, finanziellen Status, Kundenmuster und mehr zu liefern. Die SIEM-Komponente übernimmt dabei die Aufgaben der präzisen Datensammlung, des Datenmanagements und der Überprüfung von historischen Datenspeicherungen. Die **Normalisierung von Protokollen** ist ein kritischer Schritt in diesem Prozess, da SIEM sowohl Ereignisdaten als auch kontextuelle Daten sammelt. Die Normalisierung ermöglicht es, Ereignisdaten effizient in aussagekräftige Sicherheitseinblicke umzuwandeln. Dies geschieht durch das Filtern und Entfernen irrelevanter oder unerwünschter Daten, wodurch nutzlose und redundante Informationen eliminiert werden und nur relevante Daten für zukünftige Analysen beibehalten werden. Ein weiterer Aspekt ist die **genaue Bestimmung der Quellen von Protokollen**. SIEM integriert Protokolle aus verschiedenen Systemen wie Netzwerkanwendungen, Sicherheitssystemen und cloudbasierten Plattformen. Diese Komponente fokussiert sich auf die Herkunft der Daten und den Transportweg innerhalb der Organisation. Die **Hosting-Modelle** für SIEM, sei es Selbsthosting, Cloud-Hosting oder Hybrid-Hosting, bieten verschiedene Möglichkeiten für die Implementierung. SIEM nutzt die verfügbaren Protokolle, um Unregelmäßigkeiten oder bösartige Aktivitäten zu identifizieren und zu melden. Ein herausragendes Merkmal von

vgl.⁹ (Natalia G. Miloslavskaya, 2018)

vgl.¹⁰ (ARCTIC WOLF, 2023)

SIEM ist die **Echtzeitüberwachung**, die einen effektiven Schutz vor Datenverstößen ermöglicht. Diese Funktion erkennt nicht nur bösartige Angriffe, sondern lokalisiert auch präzise deren Ursprung, prognostiziert Bedrohungen und ergreift notwendige Maßnahmen, um potenzielle Datenlecks zu verhindern.

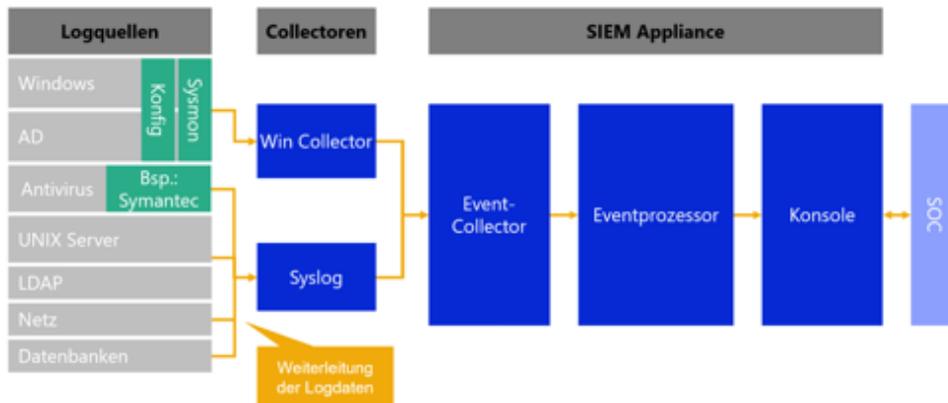


Abbildung 5: Darstellung einer SIEM-Architektur

In einem breiteren Kontext spielt SIEM eine entscheidende Rolle im Rahmen der proaktiven Sicherheitsstrategien von Unternehmen, indem es nicht nur auf Bedrohungen reagiert, sondern diese aktiv identifiziert und neutralisiert, bevor Schaden entstehen kann. Daher stellt SIEM eine zentrale Komponente für die Gewährleistung der Informationssicherheit in modernen Unternehmensumgebungen dar.¹¹

SIEM-Lösungen fokussieren sich auf die Sammlung und Analyse von Daten aus verschiedenen Sicherheits- und IT-Ereignissen. Sie bieten Funktionen wie die Erfassung, Speicherung, Standardisierung, Korrelation und Analyse von Event- und Logdaten sowie das Log-Management. Im Vergleich zu herkömmlichen SIEM-Lösungen liegt der Schwerpunkt auf erweiterten Analysen und umfassender Automatisierung, unterstützt durch Künstliche Intelligenz (KI) und insbesondere Machine Learning.

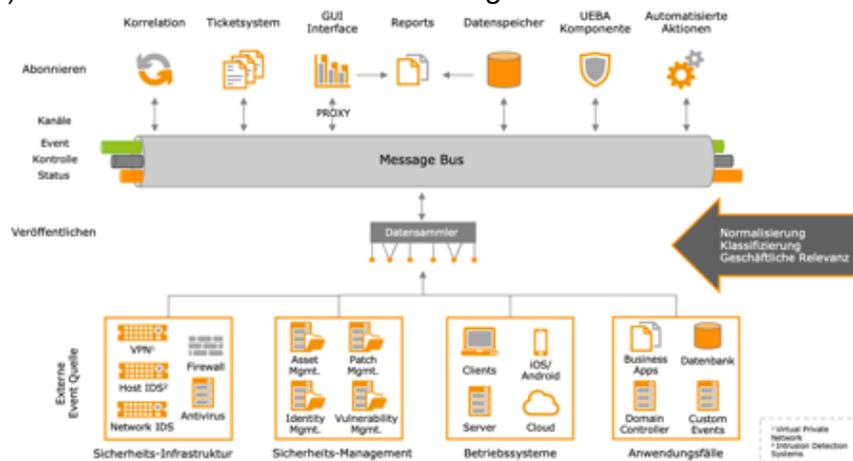


Abbildung 6: Systemkomponenten eines SIEM

Die beigefügte Grafik gibt einen Überblick über die Systemkomponenten eines SIEM. Unten sind externe Eventquellen dargestellt, die Daten aller verbundenen Systeme umfassen. Das SIEM-System speichert, standardisiert und erstellt gegebenenfalls eine Timeline. Externe Eventquellen werden von Datensammlern analysiert, klassifiziert und auf geschäftliche Relevanz überprüft, um die Daten in einen unternehmensspezifischen Kontext zu setzen.¹² Alle kommunizieren über einen Message Bus, ermöglichen "near Realtime"-Verfügbarkeit von Ereignissen und Alarmen. Die "Korrelation" nutzt ein festes Regelwerk, um Alarme zu generieren. Die "User and Entity Behavior Analytics-Komponente" (UEBA) umfasst Machine Learning-Modelle, die Abweichungen vom normalen Verhalten erkennen. Ein SIEM benötigt

vgl.¹¹ (Remya Mohanan)
vgl.¹² (KOGIT)

ein "Ticketsystem" zur systematischen Incident-Bearbeitung, kann im System oder an ein bestehendes Ticketsystem angebunden sein. Der "Datenspeicher" zentralisiert Events und Alarme, während "Automatisierte Aktionen" Identity Access Management-Systeme (IAM) oder andere direkt ansprechen können.¹³

2.1.3. SIEM Integration im Unternehmen

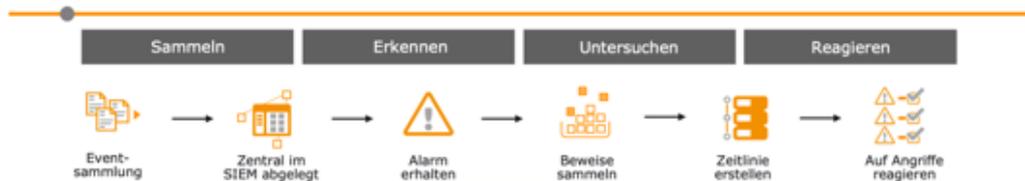


Abbildung 7: Workflow eines SIEM

Die Implementierung und Aufrechterhaltung eines effektiven SIEM-Systems erfordert die präzise Definition und professionelle Kontrolle verschiedener Betriebsprozesse, die einen reibungslosen Ablauf und die kontinuierliche Verbesserung der Sicherheitsinfrastruktur sicherstellen. Im Rahmen dieses Prozessrahmens sind mehrere Schlüsselaktivitäten identifiziert, die einen integralen Bestandteil des langfristigen Betriebskonzepts bilden.

Ein zentraler Prozess ist das **Change Management**, das eine disziplinierte Verwaltung und Dokumentation sämtlicher Systemänderungen umfasst. Dies beinhaltet die Aktualisierung von Regeln, Konfigurationen und Richtlinien, um die Integrität und Effektivität des SIEM-Systems zu gewährleisten. Die Integration neuer Systeme ist ein weiterer kritischer Schritt, der eine präzise Aufnahme von relevanten Logdaten und eine sorgfältige Konfiguration erfordert. Die **Anpassung von Sicherheitsregeln** und die kontinuierliche Definition neuer Reports ermöglichen es, auf sich wandelnde Bedrohungslandschaften und geschäftliche Anforderungen flexibel zu reagieren. Parallel dazu spielt das Security Incident Management eine entscheidende Rolle bei der Erkennung, Analyse und effektiven Bewältigung von Sicherheitsvorfällen unter Einbeziehung klar definierter Eskalationswege.

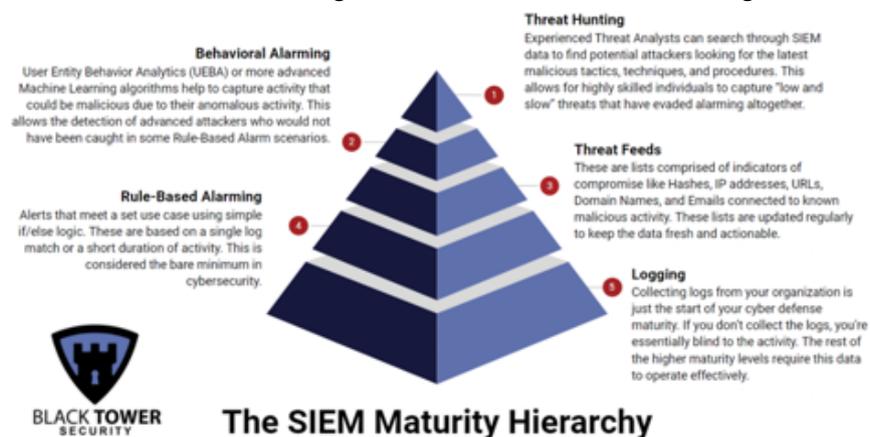


Abbildung 8: SIEM Maturity Hierachy

Die **Prozesse von Reporting und Alarmierung** gewährleisten eine zeitnahe Kommunikation und Reaktion auf sicherheitsrelevante Ereignisse. Klare Eskalationswege sind dabei unabdingbar, um eine effektive und zeitnahe Behandlung von Sicherheitsvorfällen zu gewährleisten. Die strukturierte Weiterverarbeitung von generierten Tickets und ihre langfristige Archivierung ermöglichen eine detaillierte Analyse vergangener Vorfälle. Ein weiterer essenzieller Aspekt ist das **Capacity Management**, das die Überwachung der Datenbankgröße und die Auslastung des SIEM-Systems umfasst, um eine optimale Performance sicherzustellen. Im Bereich des Security Managements erfolgt die regelmäßige Überprüfung von Reports, die Erzeugung von Sicherheitskennzahlen sowie die Anpassung von Audit Policies zur kontinuierlichen Verbesserung der Gesamtsicherheitsstruktur.

vgl.¹³ (KOGIT)

Die abschließende **Auditierung des SIEM-Systems** ist unerlässlich, um die Einhaltung von Sicherheitsrichtlinien zu überprüfen und potenzielle Schwachstellen zu identifizieren. Dabei hängt die konkrete Ausgestaltung dieser Prozesse stark ab von den individuellen Gegebenheiten, Zielsetzungen und Zuständigkeiten innerhalb eines Unternehmens. Eine flexible Implementierung dieser Betriebsprozesse ermöglicht es, das SIEM-System optimal an die spezifischen Anforderungen anzupassen und kontinuierlich zu verbessern, um so den langfristigen Erfolg und die Sicherheit der IT-Infrastruktur zu gewährleisten.¹⁴

2.1.3.1. SIEM Team mit organisatorische und technische Vorbereitungen

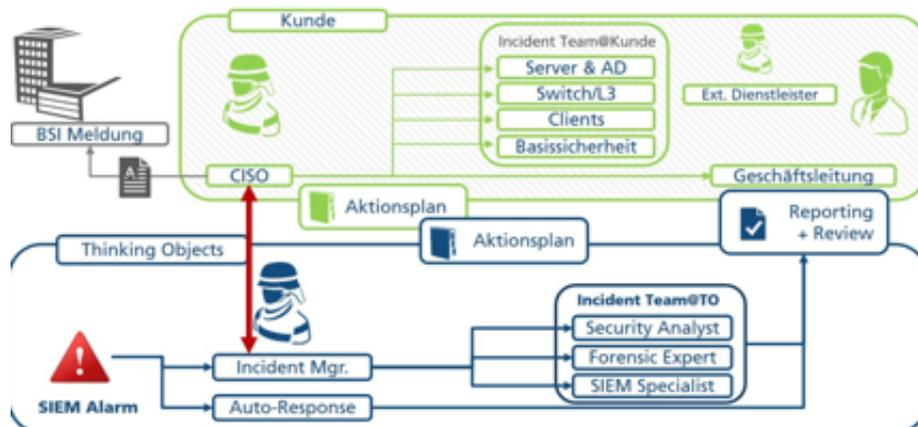


Abbildung 9: Team SIEM

Die Bildung eines **Computer Security Incident Response Team (CSIRT)** ist entscheidend für die **organisatorische Vorbereitung** zur Sicherheit der Informationstechnologie. Der CSIRT, bestehend aus Fachleuten aus verschiedenen Bereichen, koordiniert Maßnahmen von der Identifizierung bis zur Prävention von Sicherheitsvorfällen. Ein gut strukturierter CSIRT stärkt die Resilienz einer Organisation gegenüber Cyberbedrohungen und fördert das Sicherheitsbewusstsein durch Schulungen. Die erfolgreiche **technische Implementierung** eines SIEM-Systems erfordert eine sorgfältige Analyse und systematische Herangehensweise. Die Realisierbarkeitsprüfung, präzise Auswahlkriterien, allgemeine technische Voraussetzungen und die richtige Dimensionierung sind entscheidende Schritte. Die Integration verschiedener Quellformate, eine zuverlässige Netzwerktechnologie, die Aktivierung von Audit-Regeln und klare Zugriffsregelungen sind Schlüsselfaktoren für den Erfolg eines SIEM-Systems in der Sicherheitsüberwachung und -analyse.

2.1.3.2. SIEM Prozesse



Abbildung 10: SIEM-Prozesse

Im SIEM-Prozess ist die **Datensammlung** (eng. Collect data from source) der Protokolle, wobei unterschiedliche Daten und Ereignisse aus der IT-Infrastruktur einer Organisation erfasst werden. Diese Datensammlung erfolgt durch SIEM, durch eine **automatisierte**

vgl.¹⁴ (Miuccio)

Datensammlung von installierten Agenten auf den Geräten, die Protokolldateien im Syslog-Format oder Ereignisstreams sammeln. Dies ermöglicht der Organisation eine umfassende Sicht auf aktuelle Bedrohungen. Oder mit der Hilfe der **Asset- Kategorisierung** von IT-Assets wie Geräten, Netzwerken und Anwendungen. Dies ermöglicht die Überwachung von Netzwerkaktivitäten und die Identifizierung hochriskanter Assets.

Dann erfolgt die **Datenaufbereitung** (*eng.* Data Aggregation), auch als Datensammlung oder Datenerfassung bekannt. In diesem Stadium werden Sicherheitsdaten von verschiedenen Quellen innerhalb des Unternehmensnetzwerks gesammelt. Diese Quellen können Firewalls, Antivirensoftware, Netzwerkkomponenten, Serverprotokolle, Anwendungen und andere Geräte umfassen. Die Daten werden in Echtzeit oder in regelmäßigen Intervallen gesammelt, um einen ständigen Überblick über die Sicherheitslage zu gewährleisten.

Speicherung kann in große Datenmengen gesammelt werden und diese lokal, in der Cloud oder in beiden zu speichern. Hierbei sollten die Speicherorte besonders sicher sein, um Datenverlust zu verhindern. Die Datenplatzierung erfolgt nach Relevanz und Wichtigkeit sowie den aktiven Daten für die Echtzeit-Sicherheitsüberwachung sind auf Hochleistungsspeicher (in naher Zukunft mit Quantentechnology und KI) bereitgestellt, während inaktive Daten auf kostengünstigeren Speichermedien abgelegt werden. Durch die **Kategorisierung** wird durch die Optimierung und Indexierung von Daten, die Nutzung von Bedrohungsentelligenz zur Risikokategorisierung, die Anwendung von Erkennungsalgorithmen zur Reduzierung von Fehlalarmen und die Festlegung von Richtlinien für standardisierte Datenabläufe verbessert.

Nachfolgend die **Auswertung der Daten** (*eng.* Data Analysis), bei der die gesammelten Daten analysiert werden. Dieser Schritt umfasst die Verarbeitung und Interpretation der Daten, um Muster, Anomalien oder verdächtige Aktivitäten zu identifizieren und dabei werden unterschiedliche Analysetechniken eingesetzt, darunter statistische Analyse, maschinelles Lernen und Verhaltensanalyse. Die Auswertung dient dazu, potenzielle Bedrohungen zu erkennen und die Basis für die Erstellung von Regeln und Korrelationen zu legen.

Visualisierung der Daten (*eng.* Data Visualization) befasst sich mit der Darstellung der analysierten Informationen in verständlicher Form. Hierbei kommen Grafiken, Diagramme, Dashboards und Berichte zum Einsatz, um Sicherheitsanalysten und anderen Stakeholdern eine klare Übersicht über die Sicherheitslage zu bieten und helfen dabei, komplexe Zusammenhänge zu verstehen, Trends zu identifizieren und schnelle Entscheidungen zu treffen. Ein effektives Visualisierungstool ermöglicht es Sicherheitsteams, sich auf die relevantesten Informationen zu konzentrieren und schnell auf Sicherheitsvorfälle zu reagieren.

Integration von Sicherheitsinformations- und Ereignismanagementprozessen mit anderen Cybersicherheitstools schafft eine Synergie, die das gesamte Unternehmen besser schützt. Eine SIEM-Plattform, die mit verschiedenen Softwaretools integriert ist, erkennt, verhindert und begrenzt Sicherheitsbedrohungen in Echtzeit. Das Besondere an dieser Lösung ist, dass sie nicht auf bestimmte Software beschränkt ist, solange die Daten sicher auf der SIEM-Plattform integriert werden können. Beispiele für SIEM-Integrationsoptionen umfassen Software für Identitäts- und Zugriffsverwaltung, Patch-Management, Cloud-Sicherheit und Risikoverwaltung von Drittanbietern.¹⁵

2.1.3.3. SIEM Prozesse in Unternehmen durchführen

Insgesamt spielt SIEM eine Schlüsselrolle in proaktiven Sicherheitsstrategien, indem es nicht nur auf Bedrohungen reagiert, sondern diese aktiv identifiziert und neutralisiert, um potenzielle Schäden zu verhindern. Es ist eine unverzichtbare Komponente für die Informationssicherheit in modernen Unternehmensumgebungen.

vgl.¹⁵ (Mohanam)



Abbildung 11: Roadmap Erstellung SIEM Grobkonzept nach CBT Training & Consulting GmbH

Der **Kickoff-Workshop** markiert den Beginn des SIEM-Projekts und dient der klaren Definition der allgemeinen Zielsetzung. In diesem Workshop werden die grundlegenden Anwendungsfälle identifiziert, die Anforderungen präzisiert und die Zielsetzung des Projekts festgelegt. Dabei spielen auch die Rahmenparameter, wie technische Voraussetzungen und organisatorische Aspekte, eine zentrale Rolle.

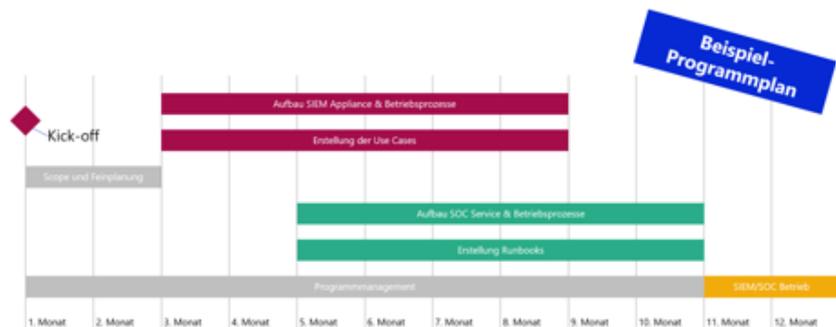


Abbildung 12: Kick-off-Programmplan

Die Ergebnisse werden sorgfältig dokumentiert, um als Leitfaden für den weiteren Verlauf des Projekts zu dienen.

Im nächsten Schritt ist die **Konkretisierung und Priorisierung** und erfolgt eine vertiefte Diskussion der **identifizierten Anwendungsfälle**. Die jeweiligen Stakeholder bringen ihre Perspektiven ein, um die Anwendungsfälle zu konkretisieren. Im Rahmen von Diskussionen im SIEM-Projektteam werden die Anwendungsfälle priorisiert, und es erfolgt eine zeitliche Planung für die Umsetzung. Die Ergebnisse dieser Phase werden ebenfalls umfassend dokumentiert.

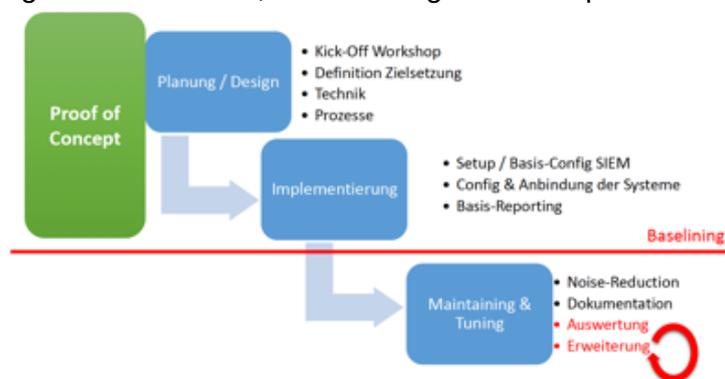


Abbildung 13: am häufigsten delegierten SOC-Anwendungsfälle

Das **Grobkonzept** wird erstellt, basierend auf den definierten Zielsetzungen, Anwendungsfällen und Rahmenparametern. Im Rahmen eines **Review- und Abstimmungs-Workshops**

wird das Grobkonzept kritisch überprüft und finalisiert. Dieser Schritt gewährleistet, dass das Konzept den Erwartungen aller Beteiligten entspricht und die Grundlage für die folgenden Umsetzungsphasen bildet.

Nach der Fertigstellung des Grobkonzepts erfolgt die **Evaluierung möglicher SIEM-Lösungen**. Dabei wird ein **Proof of Concept (PoC)** durchgeführt, um die praktische Umsetzbarkeit und Effektivität der gewählten Lösung zu überprüfen. Dieser Schritt ermöglicht es, Erfahrungen zu sammeln, potenzielle Herausforderungen frühzeitig zu erkennen und gegebenenfalls Anpassungen vorzunehmen, bevor die eigentliche Implementierung beginnt.



© Inhalte: Dipl.-Inf. Christian Brinz
© Folienmaster: CBT Training & Consulting GmbH

106

SIEM nach ISO 27001 Security Information and Event Management

Abbildung 14: Projektschritte für die Einführung von SIEMs nach CBT Training & Consulting GmbH

Ein umfassender PoC ist entscheidend für die Einführung einer SIEM-Lösung aus mehreren Gründen. Er ermöglicht die Bewertung verschiedener SIEM-Lösungen, um eine optimale Auswahl für langfristige strategische Entscheidungen zu treffen. Während des PoC können Anpassungen an technische Aspekte, Betriebsprozesse und Incident Handling vorgenommen werden, um potenzielle Herausforderungen frühzeitig zu erkennen. Der PoC sammelt konkrete Erfahrungen, um eine fundierte Entscheidung über die optimale SIEM-Lösung für die spezifischen Unternehmensanforderungen zu treffen.

Die Abstimmung mit relevanten Partnern wird durch mehr Zeit und konkrete Beispiele aus dem PoC erleichtert. Zusammengefasst ermöglicht ein ausführlicher PoC die Identifizierung der besten SIEMs-Lösung, die frühzeitige Erkennung von Herausforderungen und die aktive Einbindung aller relevanten Stakeholder im Implementierungsprozess.

Bei der **Planung und Design** im ein Kick-Off Workshop bildet den Startpunkt des SIEM-Projekts. Hier werden Ziele, technische Anforderungen und grundlegende Prozesse definiert, um das Projekt umfassend auszurichten.

In der **Implementierung** wird nach der Planung erfolgt die konkrete Implementierung des SIEM-Systems, inklusive Setup, Basis-Konfiguration und Anbindung der relevanten Systeme sowie die schrittweise Inbetriebnahme ermöglichen eine effiziente Nutzung zur Überwachung von Sicherheitsinformationen.

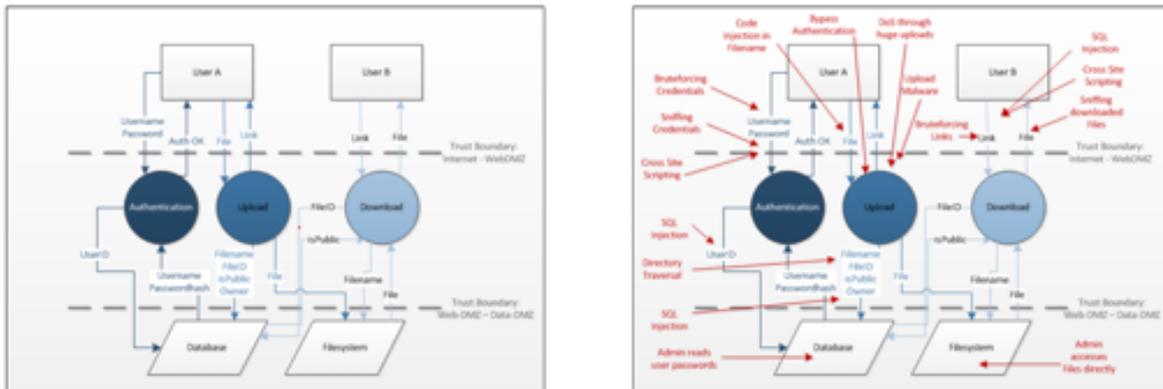
Maintaining & Tuning beinhaltet effiziente Betriebsführung und Optimierung des SIEM-Systems sind nach der Implementierung entscheidend und beinhaltet Noise-Reduction zur Minimierung von Fehlalarmen, fortlaufende Dokumentation und kontinuierliche Analyse der SIEM-Daten.

Die **Erweiterung** von SIEM-System muss langfristig an sich ändernde Anforderungen und Umgebungen angepasst werden und die Erweiterung gewährleistet, dass diese stets auf dem neuesten Stand bleiben und optimal auf die spezifischen Unternehmensanforderungen abgestimmt sind.

Die Herausforderung ist **Noise Reduction** bei der effektiven Bewältigung der ist entscheidend für die optimale Funktion des SIEM-Systems beim automatisierten Tracking von Ausnahmen, manuelle Fehlerbehebung und Filterung nach strengen Kriterien sind Schlüsselschritte in diesem Prozess.

Laufende Pflege und Erweiterung sind notwendig in einer kontinuierlichen Pflege und Erweiterung des SIEM-Systems wird durch die Herausforderung der Noise Reduction unterstrichen, um den proaktiver Ansatz gewährleisten, dass das System stets präzise und effektiv arbeitet.

Die Anwendung von **Threat Models im Monitoring** einzelner Systeme hat Vor- und Nachteile. Obwohl sie bei der Identifizierung spezifischer Bedrohungen und Schwachstellen auf Anwendungsebene helfen, ist ihre umfassende Erstellung zeitaufwendig und erfordert tiefes Fachwissen. Die Dynamik von Bedrohungsszenarien erfordert zudem kontinuierliche Anpassungen. Trotz dieser Herausforderungen bieten Threat Models klare Vorteile, da Sicherheitsteams gezielt Maßnahmen ergreifen und Schutzmechanismen implementieren können. Die umfassende Erfassung aller Bedrohungen bleibt jedoch anspruchsvoll und erfordert fortlaufendes Fachwissen. Insgesamt ist die Anwendung von Threat Models im Monitoring eine anspruchsvolle, aber effektive Methode, die kontinuierliche Anpassung erfordert, um den sich wandelnden Sicherheitsanforderungen gerecht zu werden.¹⁶



© Inhalte: Dipl.-Inf. Christian Brinz

© Folienmaster: CBT Training & Consulting GmbH

112

SIEM nach ISO 27001 Security Information and Event Management

Abbildung 15: Angriffserkennungen von SIEMs nach CBT Training & Consulting GmbH

2.1.4. SIEM Logfileanalyse

Die Herausforderungen der Logfileanalyse erfordern eine gründliche Betrachtung. Dazu zählen die Definition von Log Policies, suboptimale Standardkonfigurationen, der zeitaufwendige Export von Logs und die mangelnde Zeitsynchronisation. Organisatorische Hürden wie Datenschutzüberlegungen und Zugriffsrechte stellen weitere Herausforderungen dar. Eine durchdachte Strategie, die technische und organisatorische Aspekte berücksichtigt, ist notwendig, um die Logfileanalyse effektiv für die Sicherheitsüberwachung zu nutzen. Logfiles sind eine bedeutende Informationsquelle in der IT-Sicherheit. Sie bieten Einblicke in Systemvorgänge, sind zeitgestempelt und schwer manipulierbar. Die Praxis, Logfiles an zentrale Archive weiterzuleiten, sichert die Daten. Logfiles werden rechtlich als Beweismittel anerkannt und ihre Einfachheit ermöglicht effiziente Sicherheitsvorfallerkennung. Insgesamt sind Logfiles unverzichtbar für die moderne IT-Sicherheitsstrategie.

Diese **fünf Schritte** repräsentieren essenzielle Phasen im **Bereich der Log-Verarbeitung** innerhalb eines umfassenden Security Information and Event Management (SIEM)-Systems. Jede Phase spielt eine spezifische Rolle bei der Gewinnung von wertvollen Erkenntnissen aus den umfangreichen Datenprotokollen, die von verschiedenen Systemen und Anwendungen generiert werden:

Die **Protokollanalyse** (*eng.* Log Parsing) in dieser Anfangsphase werden Rohprotokolldaten aus unterschiedlichen Quellen gesammelt. Diese Daten können vielfältige Formate aufweisen, da verschiedene Systeme unterschiedliche Protokollstrukturen verwenden.

Das Log Parsing zielt darauf ab, diese rohen Protokolle zu extrahieren und in standardisierte, maschinenlesbare Formate umzuwandeln. Dieser Schritt schafft eine einheitliche Grundlage für die weitergehende Analyse.

Log Normalization und Kategorisierung werden die standardisierten Protokolldaten normalisiert und kategorisiert. Normalisierung bezieht sich auf die Vereinheitlichung von Protokollinformationen, um eine konsistente und vergleichbare Datenbasis zu schaffen.

vgl.¹⁶ (Infopluse)

Die Kategorisierung erfolgt durch die Zuordnung von Protokollen zu vordefinierten Typen oder Klassen. Diese Strukturierung erleichtert die spätere Identifikation von Aktivitäten und Ereignissen.

Protokollanreicherung (*eng.* Log Enrichment) beinhaltet das Hinzufügen zusätzlicher Informationen zu den standardisierten Protokolldaten. Dies können beispielsweise Geoinformationen, Benutzerkontext oder andere metadatenrelevante Details sein. Die Anreicherung verbessert die Kontextualisierung der Protokolle und ermöglicht eine präzisere Analyse.

Protokollindizierung (*eng.* Log Indexing) sind nach der Vorbereitung der Protokolldaten erfolgt die Indexierung, bei der ein effizienter Zugriff auf die Daten ermöglicht wird. Ein Index erleichtert das schnelle Auffinden von bestimmten Ereignissen oder Aktivitäten, indem relevante Schlüsselwörter oder Merkmale in einer strukturierten Form gespeichert werden. Dies beschleunigt die spätere Suche und Analyse erheblich.

Protokollspeicherung (*eng.* Log Storage) werden die finale Phase beinhaltet die langfristige Speicherung der Protokolldaten. Diese Archivierung ermöglicht retrospektive Analysen, forensische Untersuchungen und die Einhaltung von Compliance-Anforderungen.

Die Speicherung erfolgt in der Regel auf sicheren und skalierbaren Datenspeicherplattformen, um eine zuverlässige Langzeitverfügbarkeit zu gewährleisten.

Zusammen bilden diese Phasen einen kohärenten und effektiven Workflow für die Verarbeitung von Protokolldaten in einem SIEM-System. Durch die strukturierte Extraktion, Normalisierung, Anreicherung, Indexierung und Speicherung wird eine solide Grundlage für die Überwachung, Analyse und Reaktion auf Sicherheitsvorfälle geschaffen.¹⁷

2.1.4.1. Arten von Logfiles

Das **Windows Event Log** ist ein integraler Bestandteil des Windows-Betriebssystems, der Ereignisse und Aktivitäten protokolliert. Es bietet detaillierte Einblicke in Systemaktivitäten, wie Benutzeranmeldungen, Dienststart/-stopp und Firewall-Ereignisse. Das strukturierte XML-Format erleichtert die automatisierte Analyse. Der Event Viewer ermöglicht die Überwachung von Dateizugriffen.

Das **Syslog** in Unix-ähnlichen Betriebssystemen protokolliert vielfältige Ereignisse wie Benutzeranmeldungen, Systemdienste und Kernelmeldungen. Die klare Textstruktur erleichtert die automatisierte Verarbeitung, und diese ist eine wichtige Ressource für die Sicherheitsüberwachung von Linux-Systemen.

Webserver-Logs bieten Einblicke in Website-Interaktionen und ermöglichen die Analyse von Zugriffsversuchen. Das strukturierte Format erleichtert die Anomalieerkennung, während die CSV-Exportfunktion eine externe Aufzeichnung ermöglicht. Webserver-Logs sind entscheidend für die Überwachung und forensische Analyse von Webaktivitäten.

Firewall-Logs sind zentral für die Netzwerküberwachung und Sicherheit, protokollieren abgelehnte Verbindungen und bieten Einblicke in Angriffe. Ihre umfangreichen Daten erfordern sorgfältige Analyse. Firewall-Logs sind eine entscheidende Informationsquelle für die Netzwerksicherheit.

Perimeter-Geräteprotokolle von Perimeter-Geräte wie Firewalls, VPNs, IDSs und IPSs überwachen und regulieren den Netzwerkverkehr. Die Protokolle dieser Geräte sind entscheidend, um Sicherheitsereignisse zu verstehen. IT-Administratoren verwenden Syslog-Protokolle, um Sicherheitsaudits durchzuführen, operative Probleme zu beheben und den Datenverkehr im Unternehmensnetzwerk zu analysieren.

Endpoint-Protokolle wie Desktops, Laptops, Smartphones und Drucker sind Geräte, die über das Netzwerk kommunizieren. Überwachung von Endpunktprotokollen können sein auf externen Festplatten, die oft anfällig für Malware und Datenexfiltration sind sowie die Überwachung von Benutzeraktivitäten, wie Endpunktprotokolle die helfen, die Einhaltung interner und externer Richtlinien bezüglich Softwareinstallation zu überwachen. In einer zunehmend remote arbeitenden Welt stellen Endpunkte potenzielle Eintrittspunkte für bösartige Akteure dar.

vgl.¹⁷ (Exabeam)

Anwendungsprotokolle in Unternehmensanwendungen wie Datenbanken und Webserver, um Fehlerbehebung zu identifizieren und korrigieren Probleme in der Anwendungsleistung und Sicherheit und die Überwachung von Anfragen und Abfragen, helfen bei der Erkennung nicht autorisierter Zugriffe und erleichtern die Fehlerbehebung.

Internet der Dinge (IoT) bezeichnet ein Netzwerk physischer Geräte, die Daten untereinander austauschen mit Hilfe von Sensoren, Prozessoren und Software ausgestatteten Geräte ermöglichen Datensammlung, -verarbeitung und -übertragung. Wie Endpunkte generieren auch IoT-Geräte Protokolle, die Einblicke in Hardwarefunktionen und Datenflüsse bieten. Aufgrund begrenzter Speicherkapazität dieser Geräte werden die Protokolle an eine zentrale Verwaltungslösung weitergeleitet. Dort werden sie für längere Zeiträume gespeichert und von einer SIEM-Lösung analysiert, um Fehler zu beheben und Sicherheitsbedrohungen zu erkennen. Die Protokolle aus verschiedenen Quellen, einschließlich IoT, werden zentral weitergeleitet, korreliert und analysiert. Dabei kommen unterschiedliche Formate wie CSV, JSON, Key-Value-Paar und Common Event Format zum Einsatz.

Malware Scanner Logs sind entscheidend für die Sicherheitsüberwachung und -analyse. Sie enthalten Informationen über Scannerstatus, durchgeführte Scans und gefundene Malware. Die Koordination aller Scanner-Logs ist wichtig für eine umfassende Sicherheitsstrategie.

Proxy-Logs sind entscheidende Werkzeuge für die Überwachung und Analyse des Internetverkehrs. Sie enthalten detaillierte Informationen über Internetzugriffe, unterstützen die Identifikation von Missbrauch und erfordern eine verantwortungsbewusste Auswertung.

Weitere interessante Logquellen/-formate in IT-Infrastrukturen umfassen DNS-Server, E-Mail-Gateways, IDS/IPS, Applikationslogs, Middleware-Plattformen und Binary Logs von Datenbanken. Die Zusammenführung und Analyse dieser Logs ermöglichen eine umfassende Sicherheitsüberwachung, Leistungsoptimierung und Fehlerbehebung in der gesamten IT-Infrastruktur.

Die Zusammenführung und Analyse von Logs und Informationen aus diesen vielfältigen Netzwerkkomponenten ermöglicht eine umfassende Sicherheitsüberwachung, Leistungsoptimierung und Fehlerbehebung in der gesamten IT-Infrastruktur. Ein proaktives Management dieser Komponenten ist entscheidend, um eine stabile, sichere und effiziente Netzwerkinfrastruktur zu gewährleisten.^{18 19}

2.2. SIEM - Next-Generation

Next-Generation SIEMs repräsentieren eine fortschrittliche Entwicklung in der Cyber-Security-Technologie, die darauf abzielt, der zunehmenden Komplexität und Raffinesse moderner Cyberbedrohungen zu begegnen. Diese innovativen Lösungen gehen über herkömmliche SIEMs hinaus, indem sie fortschrittliche Technologien integrieren, um die Erkennung von Bedrohungen, die Incident Response und das gesamte Sicherheitsniveau zu verbessern. Eine zentrale Komponente von Next-Generation SIEMs ist die **User and Entity Behavior Analytics** (UEBA). Durch die Anwendung von künstlicher Intelligenz und Deep Learning analysieren sie Muster im Verhalten von Benutzern und Entitäten. Dies ermöglicht die Identifikation von Insider-Bedrohungen, gezielten Angriffen und betrügerischen Aktivitäten, indem normales Verhalten verstanden und Anomalien erkannt werden. Ein weiteres Merkmal ist die **Security Orchestration and Automation Response** (SOAR), die eine Integration mit Unternehmenssystemen und die Automatisierung von Incident-Response-Prozessen ermöglicht. Vergleicht man SIEMs und SOAR, so macht SIEM die Vorarbeit durch Sammlung, Korrelation, Aggregation und Analyse von Sicherheitsereignissen und -informationen aus verschiedenen Quellen, um Anomalien und Bedrohungen zu identifizieren. Hingegen SOAR sind Plattformen, die integrierte Automatisierung und Orchestrierung in die Sicherheitsreaktion, um manuelle Prozesse zu automatisieren, die auf Sicherheitsvorfällen zu folgen und sicherheitsrelevante Aktionen zu orchestrieren.

vgl.¹⁸ (Exabeam)

vgl.¹⁹ (ManageEngine)

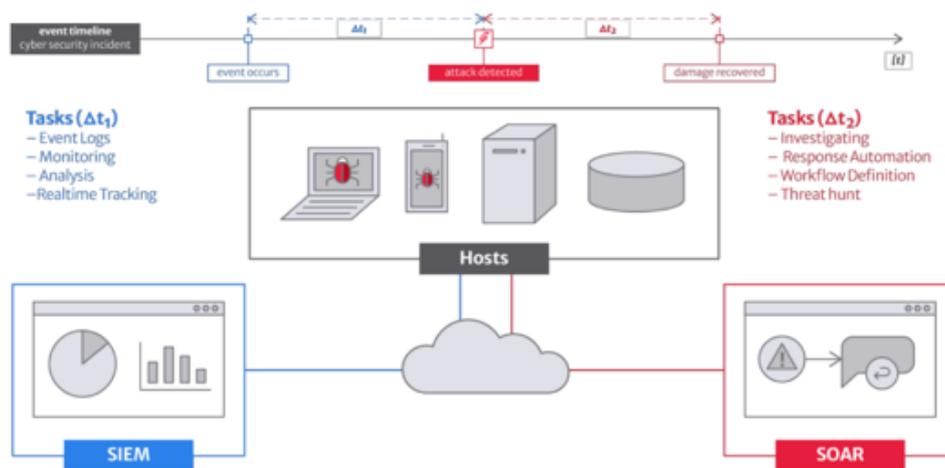


Abbildung 16: Übersicht SIEM & SOAR

Bei der Erkennung von Alarmen können Next-Gen SIEMs automatisch vordefinierte Aktionen ausführen, wie die Eindämmung einer Bedrohung oder die Benachrichtigung relevanter Stakeholder. Diese Automatisierung verbessert die Reaktionszeiten und reduziert manuelle Eingriffe. Die **Identifizierung komplexer Bedrohungen** erfolgt durch automatische Verhaltensprofilierung, wobei die Kontextanalyse von Aktivitäten genutzt wird, um Verhaltensweisen zu erkennen, die auf potenzielle Bedrohungen hinweisen. Darüber hinaus setzen Next-Gen SIEMs auf maschinelles Lernen, um Vorfälle ohne vordefinierte Regeln oder bekannte Angriffssignaturen zu identifizieren, was eine adaptive Erkennung neuartiger und raffinierter Bedrohungen ermöglicht. **Ein weiterer Schwerpunkt liegt auf der Erkennung lateraler Bewegungen** von Angreifern im Netzwerk. Durch die Analyse von Daten aus verschiedenen Netzwerkquellen können Next-Gen SIEMs solche Bewegungen identifizieren und verfolgen, was zu einem umfassenderen Verständnis des Angriffs führt. Die **Analyse des Verhaltens von Entitäten**, wie kritischen Assets im Netzwerk, ermöglicht es Next-Gen SIEMs, einzigartige Verhaltensmuster zu erlernen und automatisch Anomalien zu identifizieren. Dies trägt dazu bei, frühzeitig auf potenzielle Bedrohungen hinzuweisen. Schließlich bieten Next-Generation SIEMs eine **automatisierte Incident Response (IR)**. Sobald ein Sicherheitsvorfall erkannt wird, können diese Systeme vordefinierte Aktionen ausführen, um den Vorfall einzudämmen und zu mildern. Diese Automatisierung transformiert SIEMs in vollwertige SOAR-Werkzeuge.²⁰

Durch die Integration dieser fortschrittlichen Funktionen bieten Next-Generation SIEMs Organisationen einen robusten und proaktiven Ansatz für die Cybersecurity. Ihre Fähigkeit zur Anpassung an sich entwickelnde Bedrohungslandschaften und zur Automatisierung von Response-Prozessen macht sie zu integralen Bestandteilen moderner Cybersecurity-Strategien.

2.2.1. Herausforderung Globalisierung und technischen Systemen

Die **Globalisierung von Unternehmen** stellt im Bereich der Sicherheitsinfrastruktur, insbesondere bei SIEM-Systemen, eine spezifische Herausforderung dar. Die Implementierung und Verwaltung eines SIEMs mit weltweit verteilten Standorten erfordert eine sorgfältige Abwägung verschiedener Optionen. Eine Möglichkeit ist die zentrale SIEM-Implementierung, was zentralisierte Verwaltung und Überwachung ermöglicht. Allerdings könnten lokale Ereignisse in Niederlassungen unentdeckt bleiben. Alternativ können eigenständige SIEM-Systeme an den globalen Standorten implementiert werden, was lokale Reaktionsfähigkeit bietet, jedoch globale Korrelation erschwert. Eine dritte Option ist die Konzentration des Loggings in der Hauptniederlassung, was zentrale Speicherung ermöglicht, aber zu erhöhtem Bandbreitenbedarf führen kann. Die vierte Möglichkeit ist die Implementierung eines verteilten SIEM-Systems an den globalen Standorten, was gewisse Unabhängigkeit

vgl.²⁰ (Exabeam)

ermöglicht, jedoch höhere Lizenzkosten und Latenzen mit sich bringt. Die Wahl hängt von Faktoren wie Unternehmensgröße, globaler Präsenz, Netzwerkinfrastruktur und Sicherheits-/Compliance-Anforderungen ab. Eine umfassende Risikobewertung und Kosten-Nutzen-Analyse sind entscheidend, um eine fundierte Entscheidung zu treffen und sicherzustellen, dass die SIEM-Strategie die globalen Sicherheitsanforderungen effektiv erfüllt. Der **Einsatz von Künstlicher Intelligenz (KI) in SIEMs** ist entscheidend, um sich gegen zunehmende Cyberbedrohungen zu verteidigen. KI kann Muster in Sicherheitsdaten erkennen, Anomalien frühzeitig aufdecken und automatisierte Maßnahmen für Incident Response ermöglichen.

Durch Verhaltensanalyse lernt KI normale Benutzer- und Systemmuster, um Unregelmäßigkeiten zu identifizieren. Zudem unterstützt sie bei der Analyse von Bedrohungsinformationen, Mustererkennung und automatischer Risikobewertung. Die Integration von KI stärkt die Effizienz der SIEM-Systeme, ohne menschliche Analysten zu ersetzen, sondern ihre Arbeit zu ergänzen.

2.3. SIEM Tools

Splunk ist eine führende Plattform für das Management von maschinengenerierten Daten, insbesondere für Log-Dateien. Es ermöglicht die Suche, Überwachung und Analyse von Daten in Echtzeit. Splunk wird häufig für Sicherheitsinformationen und Ereignismanagement (SIEM) sowie für das Monitoring von IT-Infrastrukturen eingesetzt.

IBM Qradar von IBM, die sich auf die Analyse von Sicherheitsinformationen spezialisiert hat. Sie bietet Funktionen zur Ereigniskorrelation, Anomalieerkennung und Bedrohungsentelligenz. QRadar ermöglicht die zentrale Verwaltung von Sicherheitsereignissen und bietet Tools zur Untersuchung von Vorfällen.

LogRhythm werden die Sicherheitsinformationen und Ereignismanagement, Netzwerkanalyse und -überwachung sowie Endpunktsicherheit integriert. Es bietet Funktionen zur Echtzeitüberwachung, Analyse von Sicherheitsvorfällen und automatisierten Reaktionen.

Logpoint zielt ab auf umfassende Sicherheitsinformationen und Ereignismanagementfunktionen und bietet eine Plattform mit Echtzeitanalyse, Korrelation von Ereignissen, Benachrichtigungen und Tools für forensische Untersuchungen.

Rapid7 bietet verschiedene Sicherheitslösungen, darunter SIEM-Tools. Die Plattform ermöglicht die Erfassung und Analyse von Sicherheitsereignissen, Bedrohungserkennung und automatisierte Reaktionen. Rapid7 betont auch die Bedeutung von Benutzeranalysen für mehr Sicherheit.

Fusion-SIEM die sich auf die Integration von verschiedenen Sicherheitsdatenquellen konzentriert. Durch die Zusammenführung von Informationen aus verschiedenen Quellen sollen umfassendere Einblicke in die Sicherheitslage ermöglicht werden.

Graylog ist eine Open-Source-Plattform für Protokollverwaltung und Datenanalyse. Es ermöglicht die zentrale Erfassung, Indexierung und Analyse von Log-Daten. Graylog wird oft für die Überwachung von Anwendungen, Systemen und Netzwerken eingesetzt.

SolarWinds bietet eine Vielzahl von IT-Management-Tools, darunter Lösungen für Netzwerküberwachung und -verwaltung. In Bezug auf SIEM gibt es Lösungen wie SolarWinds SEM, die Ereignisprotokolle sammeln, analysieren und auf Sicherheitsbedrohungen reagieren können.

Diese Tools und weitere dienen dazu, Sicherheitsereignisse in IT-Infrastrukturen zu überwachen, analysieren und darauf zu reagieren, um die Sicherheit und Integrität von Systemen und Daten zu gewährleisten. Je nach den spezifischen Anforderungen und Präferenzen einer Organisation können unterschiedliche SIEM-Lösungen besser geeignet sein.

Einige der genannten Tools werden im nächsten Abschnitt ausführlicher beschrieben.²¹

vgl.²¹ (Tim Keary, 2024)

2.3.1. Security Monitoring Anwendungen

Die Marktdefinition und -beschreibung, wie sie von Gartner dargestellt wird, legt den Fokus auf die Perspektive zum Markt für Anwendungsleistungsüberwachung (APM) und Observability. Der Schwerpunkt liegt dabei auf transformationalen Technologien und den zukünftigen Anforderungen der Benutzer. Bei der Marktdefinition hebt Gartner hervor, dass es sich um Software handelt, die darauf ausgerichtet ist, die Leistung, Gesundheit und Benutzererfahrung von Anwendungen zu überwachen und zu analysieren. Die Zielgruppen für **APM- und Observability-Lösungen** sind vielfältig und umfassen IT-Betrieb, Site-Reliability-Ingenieure, Cloud- und Plattformteams, Entwickler sowie Produktbesitzer. Diese verschiedenen Rollen sind maßgeblich an der Verwaltung und Entwicklung von Anwendungen beteiligt. Bezüglich der **Bereitstellungsoptionen** gibt es verschiedene Möglichkeiten, darunter eigenständige Bereitstellung, vom Anbieter verwaltete gehostete Umgebung und das Software-as-a-Service (SaaS)-Modell. Die **Kernfunktionen** dieser Lösungen sind vielfältig und umfassen die Beobachtung des Transaktionsverhaltens; automatische Entdeckung und Kartierung von Anwendungen und ihrer Infrastrukturkomponenten; Überwachung über Plattformen hinweg; Identifizierung und Analyse von Leistungsproblemen; Integration mit Automatisierungs- und Servicemanagement-Tools; Geschäftsaktivitätsüberwachung; interaktive Exploration und Analyse verschiedener Telemetrietypen sowie Anwendungssicherheitsfunktionalität. Zu den **optionalen Funktionen** gehören Endpunktüberwachung für ein besseres Verständnis der Benutzererfahrung; Telemetrieaufnahme von gehosteten oder SaaS-basierten Anwendungen; AIOps-Funktionen für fortgeschrittene Analysen und Empfehlungen; Integration mit DevOps-Toolchains zur Unterstützung der kontinuierlichen oder progressiven Anwendungsbereitstellung sowie Leistungstests und Integration mit Lasttests.

Insgesamt spiegelt diese Marktdefinition die Dynamik und die vielfältigen Anforderungen wider, denen Organisationen in Bezug auf die Überwachung und Optimierung ihrer Anwendungen gegenüberstehen. Gartner orientiert sich dabei an einem zukunftsorientierten Ansatz und betont die Bedeutung von Technologien, die den sich wandelnden Benutzeranforderungen gerecht werden.



Abbildung 17: Gartner Bewertung von Tools 2023

2.3.2. IBM QRadar

Im folgenden Abschnitt 3.2.4.1. wird das Tool IBMQRadar etwas genauer beschrieben:^{22 23}

IBM positioniert sich als Herausforderer im aktuellen Magic Quadrant. Die Übernahme von Instana im Jahr 2020 unterstreicht die strategische Ausrichtung von IBM im Bereich Application Performance Monitoring (APM). Das Instana APM-Produkt wird sowohl als SaaS-Lösung als auch als selbst gehostete Option angeboten und basiert auf einer effizienten Single-Agent-Architektur. Der operative Fokus liegt vornehmlich auf Nordamerika und Westeuropa,

vgl.²² (IBM)

vgl.²³ (IBM)

wobei eine gezielte Präsenz in anderen Regionen besteht. Die Kundschaft setzt sich vorwiegend aus mittelständischen bis großen Unternehmen zusammen. Das Monitoring-Portfolio von IBM erstreckt sich über Mainframes bis hin zu modernen Cloud-Architekturen. Die zukünftige Entwicklung beinhaltet die Einführung einer synthetischen Überwachungskomponente sowie Fortschritte im Bereich KI-unterstützter Behebung von Kapazitätsproblemen.

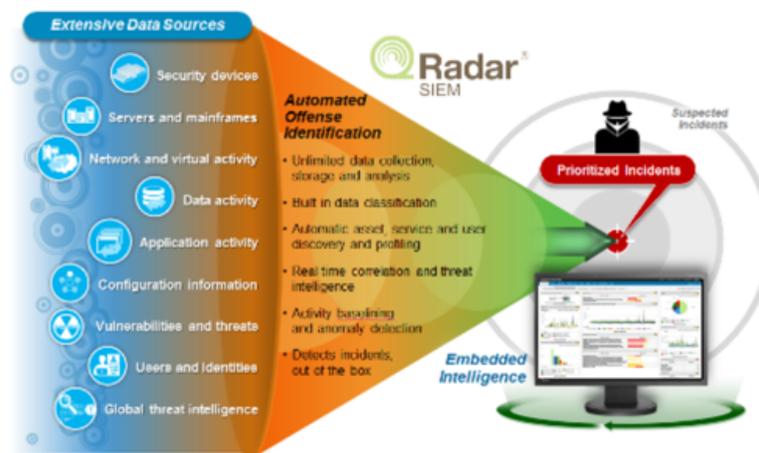


Abbildung 18: IBM QRadar

Maximieren Sie die Effizienz Ihres Sicherheitsteams mit IBM QRadar SIEM. Die Plattform minimiert wiederholende manuelle Aufgaben, ermöglicht schnelle Erkennung von Bedrohungen und verbessert Reaktionszeiten erheblich. Durch hochmoderne Funktionen wie die Integration mit SIGMA und korrelierte Protokollereignisdaten, inklusive IBM X-Force Threat Intelligence, werden zusätzliche Kontexte überflüssig. QRadar SIEM reduziert Betriebskomplexität durch nahtlose Zusammenarbeit mit Datenquellen und Sicherheitstools. Mit über 700 vorgefertigten Integrationen können Sie QRadar SIEM problemlos in vorhandene Tools integrieren, um einen umfassenden Überblick über Ihr Sicherheitsökosystem zu erhalten. Insgesamt bietet QRadar SIEM eine leistungsstarke Lösung zur Steigerung der Teamleistung, Verbesserung von Reaktionszeiten und Reduzierung der Sicherheitsinfrastrukturkomplexität.

2.3.2.1. wichtigsten Funktionen von IBM QRadar

Datenquellenintegration kann Daten aus verschiedenen Quellen sammeln, wie Protokolle von Netzwerkgeräten, Sicherheitsgeräten, Anwendungen, Betriebssystemen und mehr.

Ereigniskorrelation ermöglicht die Korrelation von Ereignissen aus verschiedenen Quellen, um komplexe Angriffsmuster zu erkennen, die möglicherweise nicht sofort ersichtlich sind.

Regelbasierte Alarmierung verwendet Regeln, um ungewöhnliche Aktivitäten oder potenzielle Sicherheitsvorfälle zu identifizieren. Wenn eine Regel ausgelöst wird, wird ein Alarm generiert.

Echtzeitüberwachung bietet Funktionen dazu an, die es Sicherheitsteams ermöglichen, Aktivitäten in Echtzeit zu verfolgen und auf Vorfälle zu reagieren.

Forensische Analyse erleichtert die forensische Analyse von Sicherheitsvorfällen, indem detaillierte Informationen über Vorfälle, Ursachen und Auswirkungen bereitgestellt wird.

Benutzer- und Entitätsanalyse (UEBA) kann auch Verhaltensanalysen auf Benutzer- und Entitätsebene durchführen, um anomales Verhalten zu identifizieren.

Integration mit Threat Intelligence kann mit Threat-Intelligence-Feeds integriert werden, um aktuelle Informationen über bekannte Bedrohungen zu erhalten und proaktiv auf diese zu reagieren.

Compliance-Management unterstützt die Einhaltung von Sicherheitsrichtlinien und -standards durch die Generierung von Berichten, die für Audits verwendet werden können.

Skalierbarkeit ist darauf ausgelegt, sich an die sich ändernde Größe und Komplexität von Netzwerken anzupassen.

Benutzeroberfläche und Dashboards bietet einer der Ersten bei der Überwachung und Analyse zu unterstützen.

Threat Investigator und Case Management werden Sicherheitsbedrohungen identifiziert. Die automatisierte Untersuchung sammelt Artefakte, führt Data-Mining durch und erstellt eine Zeitleiste des Vorfalls mit **MITRE ATT&CK-Taktiken**. Als SaaS auf AWS bereitgestellt, ermöglicht diese Methode einen schnellen Start ohne laufendes Management, damit Sie sich auf die Behebung von Sicherheitslücken konzentrieren können.

DevOps-Integration von Pipeline-Feedback-Integration ermöglicht Entwicklern eine nahtlose Integration von Observability in ihre CI/CD-Umgebung. Dies ermöglicht frühzeitige Warnungen vor Problemen bei neuen Veröffentlichungen und bietet die Möglichkeit eines Rollbacks zur Minimierung von Leistungseinbußen.

Abdeckung für Mainframes und moderne Architekturen hat erfolgreich die Überwachungsfähigkeiten auf Mainframes, insbesondere zSystems, erweitert. Darüber hinaus bietet es Lösungen für moderne containerisierte und hybride Umgebungen.

Das genaue **Preismodell** kann je nach den spezifischen Anforderungen, der Unternehmensgröße und den gewünschten Funktionen variieren. IBM bietet in der Regel flexible Lizenzierungs- und Preismodelle an, um den unterschiedlichen Bedürfnissen der Kunden gerecht zu werden, wie Lizenzierung nach Datenmenge, Module und Funktionen, Hochverfügbarkeit und Skalierbarkeit, Support und Wartung, Bereitstellungsmodelle.²⁴

Content extension type	Description
Dashboard	An associated set of dashboard items, which you view on the Dashboard tab in QRadar. Dashboard items are visual representations of saved search results.
Reports	Templates for reports that are built upon saved event or flow searches. Generate on-demand reports or schedule them to run at repeating intervals.
Saved searches	A set of search criteria (filters, time window, columns to display or group data by). By saving the criteria of commonly run searches, you don't need to define them repeatedly. Saved searches are required for reports and dashboards.
FGroup	A group of similar items by type, such as a group of log sources, a group of rules, a group of searches, or a group of report templates. FGroups are used as organizational units.
Custom rules	A set of tests that is run against events or flows that enter the system. The rule is triggered when the tests match the input. Rules can have responses, which are actions that are triggered when the conditions of a test are met. Responses can include actions such as generating an offense, generating a new event, sending an email, annotating the event, or adding data to a reference data collection.
Custom properties	Defines a property that is extracted or derived from an inbound event or flow. Can be based on a regular expression that extracts a subset of a particular event or flow payload as a textual property. They can be based on calculations, and perform an arithmetic operation on existing numeric properties of the event or flow. In QRadar V7.3.1 and later they can also be AQL functions.
Log source	A representation of a source of events such as a server, mainframe, workstation, firewall, router, application, or database. Any events that enter QRadar and originate from that source are attributed to the log source. Log sources contain the configuration information that is needed to receive inbound events, or to pull event data from the event source. Log sources contain information that is specific to your environment such as IP address or host name and other possible configuration parameters.
Log source extensions	A parsing logic definition that is used to synthesize a custom DSM for an event source for which no DSM exists. Use log source extensions to enhance or override the parsing behavior of an existing DSM.
Custom QID map entries	A combination of Event name, Event description, Severity, and Low-level category values that are used to represent a particular type of event that a log source might receive. Custom QID map entries are created to supplement the default QID map that QRadar provides for events that are not officially supported by QRadar.
Reference Data Collection	A container definition that is represented as either a set, a map, a map of sets, a map of maps, or a table for holding reference data. Searches and rules can refer to Reference data collections.
Historical Correlation Profile	A combination of a saved search and a set of one or more rules. Use historical correlation profiles to test rules by rerunning a set of historical events through an offline version of the custom rule engine that has a subset of rules enabled.
Custom Functions	An SQL-like function (defined in JavaScript) that you can use in an Advanced search to enhance or manipulate data.
Custom Actions	A custom response for a rule to run, when the rule is triggered. Custom actions are defined by a Python, Perl, or Bash script that can accept arguments from the event or flow data that triggered the rule.
Building Block	A group of commonly used tests to build complex logic so that they can be used in rules. Building blocks use the same tests that rules use, but have no actions that are associated with them, and are often configured to test groups of IP addresses, privileged user names, or collections of event names.

Abbildung 19: IBM QRadar Types of content extensions

2.3.2.2. IBM QRadar Suite

IBM® Security QRadar EDR bietet Sicherheitsanalysten einen umfassenden Einblick in das Endpunktökosystem. Nahtlos in QRadar SIEM integrierbar, ermöglicht die Lösung die effektive Reaktion auf Endpunktbedrohungen und verbessert die Gesamtsicherheitsinfrastruktur.

QRadar Log Insights erleichtert Sicherheitsanalysten die Arbeit durch eine cloudnative Lösung für das Protokollmanagement und die Beobachtbarkeit von Sicherheitsvorgängen. Sie bewältigt mühelos die Protokoll-Datenlast eines gesamten Unternehmens und ermöglicht schnelle Reaktionen auf Sicherheitsereignisse.

vgl.²⁴ (IBM, 2024)

QRadar SOAR koordiniert und automatisiert Reaktionen auf SIEM-Warnungen, indem es handlungsrelevante Informationen bereitstellt. Die nahtlose Integration in die QRadar-Umgebung ermöglicht effiziente Zusammenarbeit von Sicherheitsteams bei der Bewältigung von Sicherheitsvorfällen.

IBM® Security QRadar SIEM bietet eine umfassende Lösung für die Sicherheitsüberwachung, sowohl in der Cloud als auch on Premises. Es ermöglicht Unternehmen, Daten und Sicherheitsanalysen zu überwachen, um schnell auf kritische Bedrohungen zu reagieren und die Sicherheitslage zu stärken.²⁵

2.3.2.3. IBM QRadar Suite mit KI

Die **Ariel Query Language (AQL)**, extrahiert Ereignis- und Flussdaten, filtert und führt aus Aktionen, die aus der Ariel-Datenbank in IBM® QRadar® extrahiert wird sowie Daten abzurufen, auf die möglicherweise nicht einfach über die Benutzeroberfläche zugegriffen werden können.²⁶

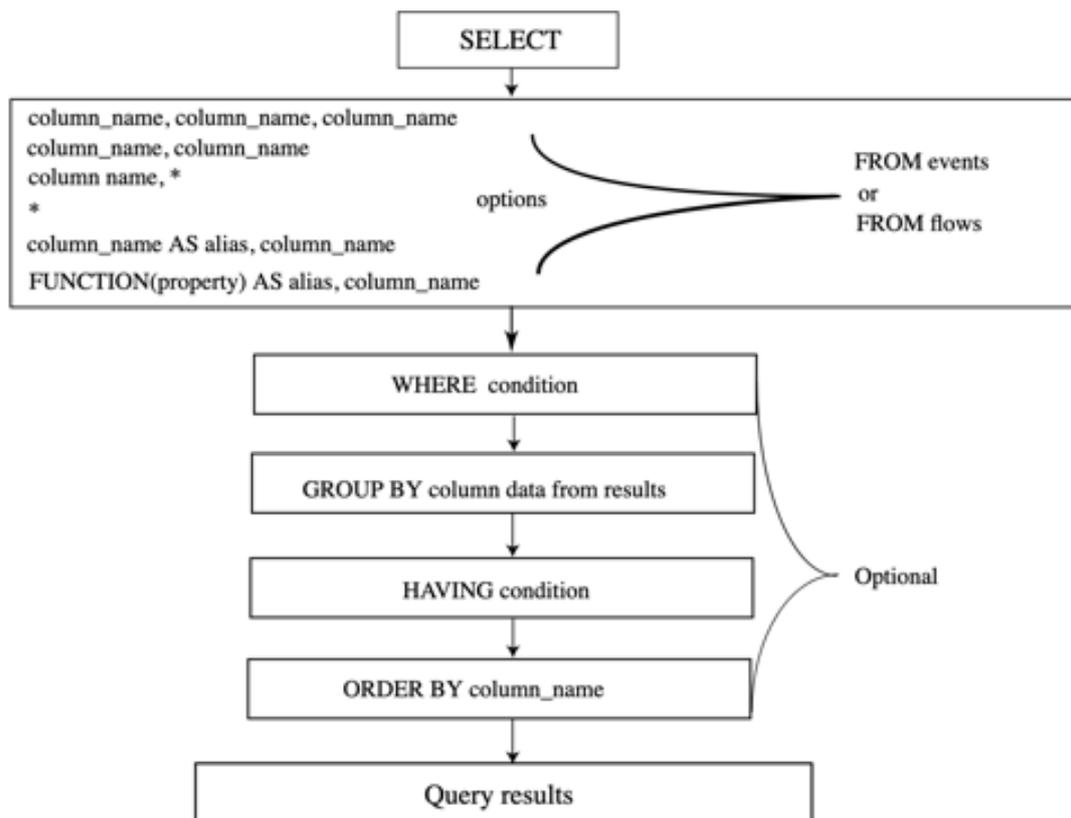


Abbildung 20: AQL query flow

vgl.²⁵ (IBM)
vgl.²⁶ (IBM, 2023)

Basic AQL Commands	Comments
<code>SELECT * FROM events LAST 10 MINUTES</code>	Returns all the fields from the events table that were sent in the last 10 minutes.
<code>SELECT sourceip,destinationip FROM events LAST 24 HOURS</code>	Returns the sourceip and destinationip from the events table that were sent in the last 24 hours.
<code>SELECT * FROM events START '2017 01 01 9:00:00' STOP '2017 01 01 10:20:00'</code>	Returns all the fields from the events table during that time interval.
<code>SELECT * FROM events limit 5 LAST 24 HOURS</code>	Returns all the fields in the events table during the last 24 hours, with output limited to five results.
<code>SELECT * FROM events ORDER BY magnitude DESC LAST 24 HOURS</code>	Returns all the fields in the events table sent in the last 24 hours, sorting the output from highest to lowest magnitude.
<code>SELECT * FROM events WHERE magnitude >= 3 LAST 24 HOURS</code>	Returns all the fields in the events table that have a magnitude that is less than three from the last 24 hours.
<code>SELECT * FROM events WHERE sourceip = '192.0.2.0' AND destinationip = '190.51.100.0' START '2017 01 01 9:00:00' STOP '2017 01 01 10:20:00'</code>	Returns all the fields in the events table that have the specified source IP and destination IP within the specified time period.
<code>SELECT * FROM events WHERE INCIDR('192.0.2.0/24', sourceip)</code>	Returns all the fields in the events table where the source IP address is within the specified CIDR IP range.
<code>SELECT * FROM events WHERE username LIKE 'brou%'</code>	Returns all the fields in the events table where the user name contains the example string. The percentage symbols (%) indicate that the user name can match a string of zero or more characters.
<code>SELECT * FROM events WHERE username ILIKE 'brou%'</code>	Returns all the fields in the events table where the user name contains the example string, and the results are case-insensitive. The percentage symbols (%) indicate that the user name can match a string of zero or more characters.
<code>SELECT sourceip,category,credibility FROM events WHERE (severity > 3 AND category = 5018)OR (severity < 3 AND credibility > 8)</code>	Returns the sourceip, category, and credibility fields from the events table with specific severity levels, a specific category, and a specific credibility level. The AND clause allows for multiple strings of types of results that you want to have.
<code>SELECT * FROM events WHERE TEXT SEARCH 'firewall'</code>	Returns all the fields from the events table that have the specified text in the output.
<code>SELECT * FROM events WHERE username ISNOT NULL</code>	Returns all the fields in the events table where the username value is not null.

Abbildung 21: Simple AQL queries

Bei der Planung oder Erstellung einer IBM® QRadar®-Bereitstellung ist es wichtig, die QRadar-Architektur zu verstehen.

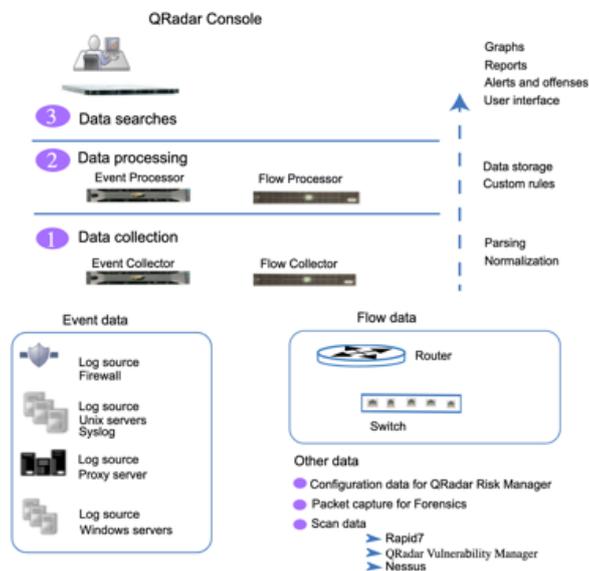


Abbildung 22: QRadar Architektur

QRadar sammelt, verarbeitet und speichert Echtzeitnetzwerkdaten zur Netzwerksicherheitsverwaltung. Die modulare Architektur von QRadar bietet Echtzeit-Sichtbarkeit der IT-Infrastruktur für Bedrohungserkennung und -priorisierung. Die Betriebsstruktur besteht aus drei Ebenen: Datensammlung, Datenverarbeitung und Datenabfragen. Die Architektur funktioniert unabhängig von der Bereitstellungsgröße und umfasst Funktionen wie Ereignis- und Flussdatensammlung, benutzerdefinierte Regelerstellung, Bedrohungsabwehr und forensische Untersuchungen. Die Benutzer können über die QRadar-Konsole auf gesammelte und verarbeitete Daten zugreifen, um Suchvorgänge, Analysen und Berichte durchzuführen sowie Warnungen und Verstöße zu untersuchen.²⁷

IBM Z® Anomaly Analytics enthält das Z-Topology Insight Pack, das eine Anomaliebewertung des metrikbasierten ML-Systems im Zusammenhang mit einer entdeckten Topologie der z/OS®-Rechenzentrumsressourcen zu visualisieren kann. Um das Z-Topology Insight Pack verwenden zu können wird das IBM Z-Resource Discovery Data Service-Modul mit einem Problem Insights-Server zu integriert, das auf einem verwendeten Linux®-Betriebssystem installiert wird, wie in Systemanforderungen für die Softwarecontainer angegeben werden. Dabei wird das Z-Topology Insight Pack nicht vom Problem Insights-Server auf IBM® z/OS UNIX System Services unterstützt wird und auf die GUI zuzugreifen zu können, wählt man für eine metrikbasierte Anomalien ein Sysplex, ein System und ein Subsystem aus und dann auf "Explore Topology" zu klicken.^{28 29}

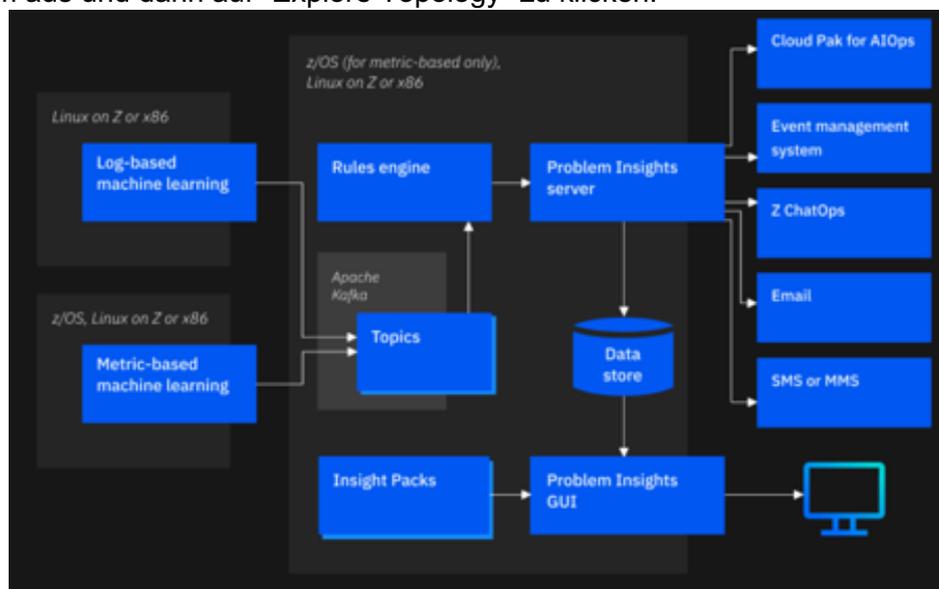


Abbildung 23: Problem insights overview

Der Problem Insights-Server in IBM Z® Anomaly Analytics bietet Einblicke in potenzielle Probleme in Ihrer IT-Umgebung und ermöglicht das Durchsuchen von Mustern und Trends anomalen Verhaltens für überwachte z/OS®-Subsysteme. Er kann auch mit Ereignismanagement-Systemen integriert werden, um Warnungen für erkannte Ereignisse zu verwalten. Der Problem Insights-Server ist ein flexibles Anwendungsframework, das verschiedene Betriebsanalyse-Anwendungsfälle unterstützt und kann durch Mikroservices und Web-UI-Einblicke erweitert werden und dazu dient Analyseergebnisse zu visualisieren. Dabei sollte der Problem Insights-Server auf z/OS UNIX System Services (unterstützt nur metrik-basierte maschinelle Lerneinblicke) oder ein Linux® auf Z- bzw. auf ein x86_64/amd64 System installiert sein.

1. Der Datenfluss in IBM Z-Anomaly Analytics zeigt, dass der Problem Insights-Server durch ML-Funktionen in Echtzeit- und historische Analyseergebnisse aufruft und diese mit den Datenspeichern interagieren und visualisieren.

vgl.²⁷ (IBM, 2023)

vgl.²⁸ (IBM, 2024)

vgl.²⁹ (IBM, 2024)

2. Durch Erkennung der Analyseergebnisse werden die Regeln-Engine-Mirco-Service, die vorgegeben werden von User, überwacht und die Scoring-Ergebnisse vorübergehend im Apache Kafka-Nachrichtenbroker gespeichert, um Anomalieereignisse zu erkennen.
3. Die gefundenen Anomalieereignisse, die den Regeln entsprechen, werden im Datenergebnissatz im Datenspeicher erzeugt und visuell im Problem Insights GUI angezeigt.
4. Optional können Anomalieereignisse an bis zu vier verschiedene Ereignisziele weitergeleitet werden, wie mehrere Ereignismanagement-Systeme, IBM Z ChatOps, EMail-, Short Message Service (SMS)- und Multimedia Messaging Service (MMS)-Ziele.³⁰

In folgendem wird der **Datenfluss** zwischen dem Protokoll basierenden ML in IBM Z Anomaly Analytics auf IBM z/OS UNIX System Services beschrieben:³¹

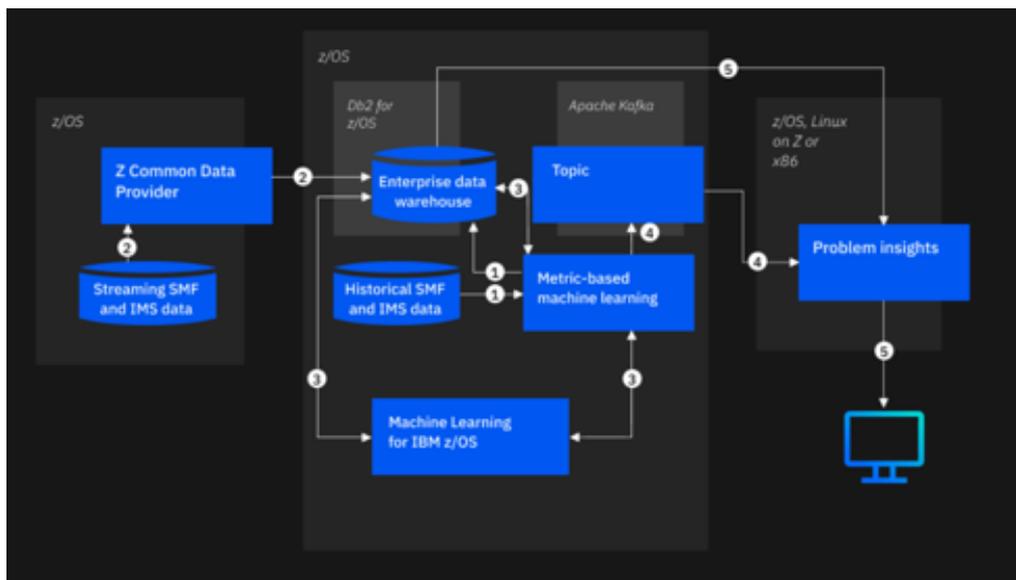


Abbildung 24: Metric-based machine learning on z/OS

1. Historische Daten werden im Batch-Modus geladen.
2. SMF- und IMS-Daten werden über den Z-Common Data Provider ins IBM Db2 for z/OS Enterprise Data Warehouse verarbeitet.
3. Der Z-Common Data Provider sammelt Daten in Echtzeit und sendet sie an das Data Warehouse.
4. Ein metrikbasiertes maschinelles Lernmodell vergleicht Daten mit dem Modell, identifiziert Abweichungen und speichert Ergebnisse im Data Warehouse.
5. Anomaliebewertungen gehen an Apache Kafka.
6. Die Regeln-Engine überwacht Apache Kafka und zeigt stark anomale Ereignisse im GUI an.
7. Der Problem Insights Server holt Bewertungsergebnisse aus dem Data Warehouse und zeigt sie im GUI an.
8. Auf Linux kann das IBM Z Resource Discovery Data Service-Modul mit dem Problem Insights Server integriert werden, um Anomaliebewertungen im Kontext der z/OS-Ressourcentopologie zu visualisieren.

vgl.³⁰ (IBM, 2024)

vgl.³¹ (IBM, 2024)

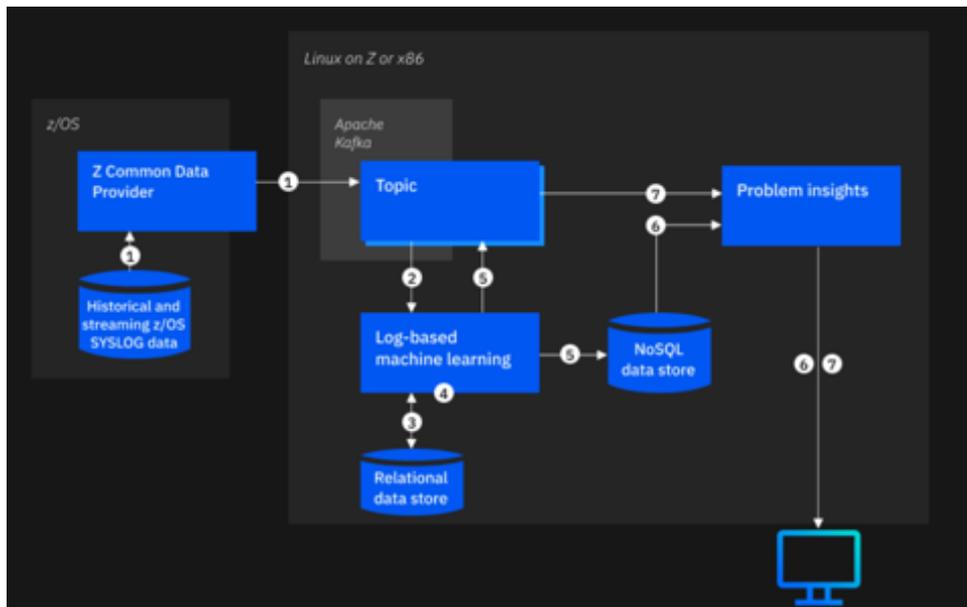


Abbildung 25: Log-based machine learning on Linux

In folgendem wird der **Datenfluss** zwischen dem Protokoll basierenden maschinellen Lernsystems in IBM Z-Anomaly Analytics auf Services Linux System Services beschrieben:³²

1. Historische oder Echtzeit-Z/OS SYSLOG- oder OPERLOG-Daten werden vom Z-Common Data Provider gesammelt und an Apache Kafka weitergeleitet.
2. Das auf Protokoll basierende maschinelle Lernen abonniert das Apache Kafka SYSLOG-Thema und analysiert die Daten.
3. Eingehende Protokoll Daten werden zusammengefasst und in einer relationalen Datenbank gespeichert.
4. Falls ein Protokoll-Datenmodell vorhanden ist, wird die Zusammenfassung mit dem Modell verglichen, um abnormales Verhalten zu identifizieren.
5. Die Ergebnisse werden in einem NoSQL-Langzeitdatenspeicher gespeichert und an den Apache Kafka-Broker geschrieben.
6. Wenn Benutzer auf die Problem Insights-GUI zugreifen, werden die Ergebnisse aus dem NoSQL-Datenspeicher abgerufen und in der GUI dargestellt.
7. Die Regel-Engine von Problem Insights überprüft Apache Kafka nach anomalen Ereignissen. Hoch anomale Ereignisse werden in der GUI angezeigt und können an maximal vier Eventziele weitergeleitet werden.

Das **Insight Pack für Protokollanomalieerkennung** beruht auf dem Problem das Insights-Server, indem das Insight Pack für Protokollanomalieerkennung aktiviert sein müssen, um subsystemspezifische und spezifischer Erkenntnisse zu erhalten sowie Einblicke visuell darstellen zu können, die unter der Installation oder Aktualisierung von Insight-Paketen unter Linux/UNIX sich befinden. Das Problem der Insights GUI enthält metrische Anomalien, wie die Registerkarte "Metrikbasierte Anomalien" in tabellarischer Ansicht darstellt, die Anomalieinformationen für alle aktivierten Subsysteme im metrikbasierten ML enthält. Diese Optionen zeigen das Zeitfenster an und legen die Anzeige der Anomalieinformationen fest; herausfiltern von Anomalieinformationen nach Sysplex und Subsystemtyp filtern; zugreifen direkt auf subsystemspezifische Scorecards (normal 0-39 Score, niedrig 40-89, 90-100); das mehrere Subsysteme werden verglichen, um anomale Aktivitäten zu korrelieren oder mit dem optionalen Z-Topology Insight Pack in eine Topologieansicht wechseln.

Das **Problem Insights GUI für Protokollanomalieerkennung** erfolgt durch Klicken auf die Registerkarte "Protokollbasierte Anomalien", wo das Problem Insights-GUI angezeigt wird:

vgl.³² (IBM, 2024)

Analysen erfolgen durch die Auswahl des Datums und der zu überwachenden Systeme. Anzeige der höchsten Anomaliepunktzahl für jede überwachte Stunde.

Nachrichtenverlauf erstellt eine Anzeige aller Vorkommen einer Nachricht über alle überwachten Systeme hinweg.

Benachrichtigungen enthalten Nachrichten aus protokollbasierten ML über Aktivitäten im System, die Beachtung erfordern und auf den Erfolg oder Nichterfolg der Trainingsdaten hinweisen.

Systemstatus zeigt Statusinformationen für überwachte Systeme an, inbegriffen den Status der Datenpipeline aller Komponenten bei der Protokollverarbeitung durch protokollbasierte ML-Systeme.

Verwaltung Trainingsmanagement ist eine Verwaltung von Trainingsstatus und Details für überwachte z/OS-Systeme, einschließlich der Möglichkeit, Nachrichten zu ignorieren, Training anzufordern und Trainingsdaten zu überprüfen, dabei werden z/OS-Systemnachrichten ignoriert oder wiederhergestellt, Anzeige einer Trainingsanfrage angefordert oder angezeigt, Trainingsdatenmodell aktualisiert oder angezeigt.

Modellbeginndatum sind aktuelle Modelldaten, Modelltrainingsdaten, mit Beginn und Enddatum, Trainingszeiträume.

Administration und Konfiguration von Trainingszeiträumen, -intervallen könne konfiguriert werden und sich auf das für ein überwachtes z/OS-System auswirken.

Variablenanalyse von Variablen in z/OS SYSLOG-Nachrichten ist nun möglich. Durch Aktivierung der Variablenanalyse extrahiert das auf Protokollen basierende maschinelle Lernen Variablen aus Nachrichten und bewertet ihre Seltenheit im Vergleich zu anderen Nachrichten. Dabei werden historische Muster genutzt und in einem Modell gespeichert.

Das System erstellt zunächst ein erstes Modell, das täglich mit **maximal 15 Tagen an z/OS SYSLOG-Meldungen aktualisiert** wird. Diese Meldungen werden dann in der Problem Insights-GUI mit dem ihnen zugewiesenen Seltenheitswert angezeigt. Auf einer **Skala von 0 bis 1 steht 0 für eine häufige Variable, während 1 für sehr selten** steht. „**Not available**“ wird für Meldungen verwendet, die nicht bewertet werden können, z. B. wenn die Meldung keine Variablen enthält oder wenn sie zwar Variablen enthält, aber nicht genügend Informationen für eine Bewertung vorhanden sind.³³

Die **IBM® watsonX™-Plattform** bietet eine organisatorische Ansicht, sodass jeder Benutzer problemlos seine maßgeschneiderte KI-Lösung einrichten, alle Datenquellen effektiv verwalten und verantwortungsvolle KI-Workflows beschleunigen kann. Die Plattform umfasst drei Kernkomponenten und eine Vielzahl von KI-Assistenten, die auf das Ziel hinarbeiten, vertrauenswürdige Daten im gesamten Unternehmen zu skalieren und das Beste aus KI herauszuholen. Watsonx nutzt eine offene, modulare Technologiearchitektur, die eine Vielzahl von Modellen für unterschiedliche Geschäftsanwendungsfälle und Compliance-Anforderungen unterstützen kann. Diese wiederum helfen dabei, sich auf unternehmensrelevante Bereiche wie Personalwesen, Kundendienst oder IT-Betrieb zu konzentrieren, um neue Möglichkeiten zur Wertschöpfung zu erschließen.

Die Plattform legt großen Wert auf Vertrauenswürdigkeit und wurde unter Berücksichtigung der Grundsätze von Transparenz, Rechenschaftspflicht und Governance entwickelt. Dies ermöglicht die Berücksichtigung rechtlicher, regulatorischer, ethischer und Genauigkeitsaspekte im Lichte der aktuellen KI-Vorschriften. Mit watsonx verlagert sich die Rolle des Nutzers von der bloßen Bedienung von KI hin zur strategischen Nutzung von KI als Quelle der Wertschöpfung – Modelle, die einen direkten Mehrwert für das Unternehmen und seine Geschäftsprozesse generieren.³⁴

vgl.³³ (IBM, 2024)

vgl.³⁴ (IBM)

2.3.2.4. Demo von IBM QRadar

IBM Security QRadar SIEM Test-Demo unter [Book a demo](#) zu buchen.³⁵

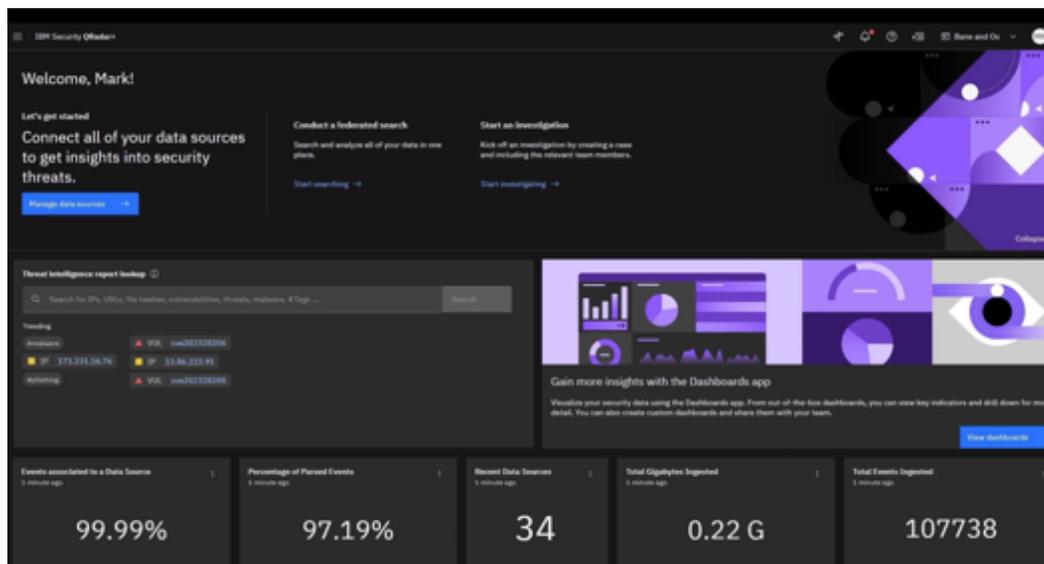


Abbildung 26: IBM Security QRadar SIEM Demo

2.3.3. Logpoint

Im folgenden Abschnitt 3.2.4.2. wird das Tool Logpoint etwas genauer beschrieben.³⁶

Logpoint bietet eine integrierte SIEM-Lösung, die nativ mit SOAR-Funktionalitäten kombiniert ist. Die Nutzung dieser Lösung erfolgt hauptsächlich in SaaS- oder Cloud-basierten Umgebungen, wobei der primäre Zielmarkt kleine und mittlere Unternehmen, vor allem in Europa, sind. Besonders hervorzuheben ist Logpoints starke Fähigkeit, Bedrohungen in SAP-Umgebungen zu überwachen und einzudämmen. Die Lizenz der SIEM-Lösung basiert auf der Anzahl der überwachten „Knoten“. Die SOAR-Komponente wird nach der Anzahl der Benutzer berechnet und eine Einzelbenutzerlizenz ist bereits im SIEM-Produkt enthalten. UEBA-Funktionalitäten werden separat lizenziert und ergänzen das breite Sicherheitsangebot von Logpoint.



Abbildung 27: Logpoint Plattformen

³⁵ (IBM)
vgl. ³⁶ (Logpoint)

2.3.3.1. wichtige Funktionen von Logpoint

Die nachfolgenden Funktionen sind laut Logpoint "Die Top 10 SIEM-Use-Cases für die Implementierung" und werden kurz beschrieben Code-Beispielen von Logpoint:³⁷

"Erkennung kompromittierter Benutzer-Anmeldedaten um sicherstellen, dass klare Anwendungsfälle und Arbeitsabläufe existieren, wie Brute Force, Pass-the-Hash, Golden Ticket oder andere und diese zu kompromittieren. Falls dies der Fall ist, erfolgt eine Identifikation der betroffenen Benutzer und Einrichtungen, um weiteren Schaden zu vermindern.

```
[label=User label=Login label=Successful logon_type=9 package=Negotiate logon_process=seclogon] as s1 followed by [norm_id=WindowsSystemon label="Process" label=Access image="*lsass.exe" access="0x1010"] as s2 within 5 second on s1.host=s2.host| rename s1.host as host, s1.user as user, s1.targetoutboundusername as target_user, s2.process as "process"| chart count() by user, target_user
```



Abbildung 28: Erkennung kompromittierter Benutzer-Anmeldedaten

Nachverfolgung von Systemänderungen richtet System Change Tracking geeignete Regeln ein, um kritische Ereignisse wie unbefugte Konfigurationsänderungen oder Löschungen von Audit-Trails zu kennzeichnen, um Schäden zu verhindern und zusätzliche Risiken wie die Manipulation von Audit-Protokollen zu minimieren.

```
label=Log label=Clear | chart count() by log_ts, host, user, target_channel
```



Abbildung 29: Nachverfolgung von Systemänderungen

Erkennung von ungewöhnlichem Verhalten bei privilegierten Konten wie System- oder Datenbankadministratoren mit speziellen Zugriffsrechten und genaue Überwachungen bzw. Durchsuchung von ungewöhnlichem Verhaltenmustern, da diese auf Bedrohung oder Kompromittierung enthalten könnten.

```
[norm_id=WinServer label=User label=Login label=Remote host IN WINDOWS_DC user IN ADMINS] as s1 followed by [norm_id=WinServer label=User label=Login label=Remote -host IN WINDOWS_DC] as s2 within 10 minute on s1.user=s2.user | rename s1.host as domain_controller, s2.host as host, s1.user as user, s1.domain as domain | chart count() by domain_controller, host, user, domain
```

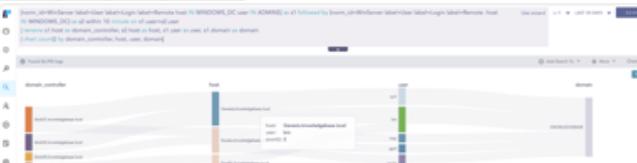


Abbildung 30: Erkennung von ungewöhnlichem Verhalten bei privilegierten Konten

Cloudbasierte Anwendungssicherheit berücksichtigt Cloud Computing zur Erfüllung von Compliance-Anforderungen, zur Verbesserung von Benutzerüberwachung und Zugriffskontrolle sowie von Verhinderung von potenzielle Malware-Infektionen und Datenlecks, wie die cloudbasierten Anwendungen in Salesforce, Office365 oder AWS sollten als Protokolldatenquellen unterstützt werden.

```
norm_id IN CLOUD_APPLICATIONS | chart count() by norm_id
```

vgl.³⁷ (Logpoint)



Abbildung 31: Sicherheit für cloudbasierte Anwendungen

Erkennung von Phishing-Angriffen (z.B. schnelle Spear-Phishing), finden schnell Schwachstellen beim Netzwerkzugriff und ermöglichen Analysten, sofort Maßnahmen zu ergreifen, indem sie nachverfolgen, wer Phishing-E-Mails empfängt, anklickt oder darauf antwortet.

`label=Email label=Receive subject IN ['Payroll Deduction Form', 'Please review the leave law requirements', 'Password Check Required Immediately', 'Required to read or complete: "COVID-19 Safety Policy"', 'COVID-19 Remote Work Policy Update', 'Vacation Policy Update', 'Scheduled Server Maintenance -- No Internet Access', 'Your team shared "COVID 19 Amendment and Emergency leave pay policy" with you via OneDrive', 'Official Quarantine Notice', 'COVID-19: Return To Work Guidelines and Requirements'] | chart count() by source_address, subject, sender, receiver, destination_host`

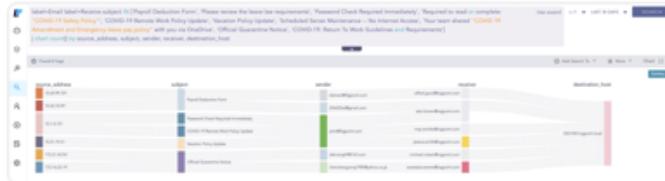


Abbildung 32: Erkennung von Phishing-Angriffen

Überwachung von Auslastung und Verfügbarkeit erfolgt indem die Auslastung, Verfügbarkeit und Antwortzeiten verschiedener Server und Dienste mithilfe geeigneter Korrelationsregeln und Warnmeldungen kontinuierlich überwacht werden, um Ausfälle und Überlastungen frühzeitig zu erkennen und Ausfallzeiten und damit verbundene Kosten zu vermeiden.

`label=Memory label=Usage | timechart max(use) as memoryUsePercent every 1 hour`



Abbildung 33: Überwachung von Auslastung und Verfügbarkeit

Logdaten-Management ermöglicht grosse Datensätze, wie Logdaten, von Datenbanken, Anwendungen, Benutzern und Server zu normalisieren und zu zentralisieren, um eine Erzeugung von nahtloser Durchführung von Analysen und Korrelationen sicherheitsrelevanter Ereignisse zu erstellen und ermöglicht damit das IT-Sicherheitsteam, Daten nach bestimmten Metadaten oder Werten zu durchsuchen.

`label=Access label=Object`



Abbildung 34: Logdaten-Management

SIEM für General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA) oder PCI-Compliance werden erfüllt von verschiedene Compliance-Vorschriften, die im SIEM-System dokumentiert mit wann und von wem auf welche Daten zugegriffen werden und unterstützt somit die Compliance-Anforderungen zu erfüllen und Verstöße zu verhindern unterstützen.

`label=File label=Modify | chart count() by log_ts, user, user_type, source_address, domain, file, application`

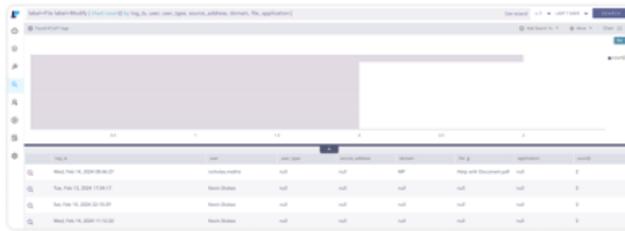


Abbildung 35: SIEM für GDPR, HIPAA oder PCI-Compliance

Threat Hunting (Erkennung nach Bedrohungen) wird aktiv genutzt, um Cyberrisiken im Unternehmen oder Netzwerk proaktiv zu identifizieren und um auf schnelle Reaktion von Sicherheitsproblemen und neue, unbekannte Bedrohungen Angriffe oder Sicherheitslücken zu erkennen und zu reagieren.

```
norm_id=WindowsSysmon event_id=11 file="*.exe" path IN [
"C:\ProgramData*", "**\AppData\Local*", "**\AppData\Roaming*",
"C:\Users\Public*"] -source_image IN ["**\Microsoft Visual Studio\Installer*\BackgroundDownload.exe",
"C:\Windows\system32\cleanmgr.exe", "**\Microsoft\Windows Defender\MsMpEng.exe",
"C:\Windows\SysWOW64\OneDriveSetup.exe", "**\AppData\Local\Microsoft\OneDrive*", "**\Microsoft\Windows
Defender\platform\MpCmdRun.exe", "**\AppData\Local\Temp\mpam-*.exe*"] chart count() by host, file, source_image
```



Abbildung 36: Suche nach Bedrohungen (Threat Hunting)

SIEM für die Automatisierung von Bedrohungserkennung erstellt worden und ermöglicht automatische Reaktionen auf Sicherheitsvorfälle, leitet Warnungen an LogPoint SOAR weiter und beschleunigt die Reaktion auf Vorfälle durch die Automatisierung manueller Aufgaben, steigert die SOC-Produktivität und reduziert Kosten. Dabei ist die LogPoint SOAR kostenlos in Ihrer SIEM-Lizenz für die gesamte IT-Infrastruktur zu erhalten.³⁸



Abbildung 37: SIEM für die Automatisierung

vgl.³⁸ (Logpoint)

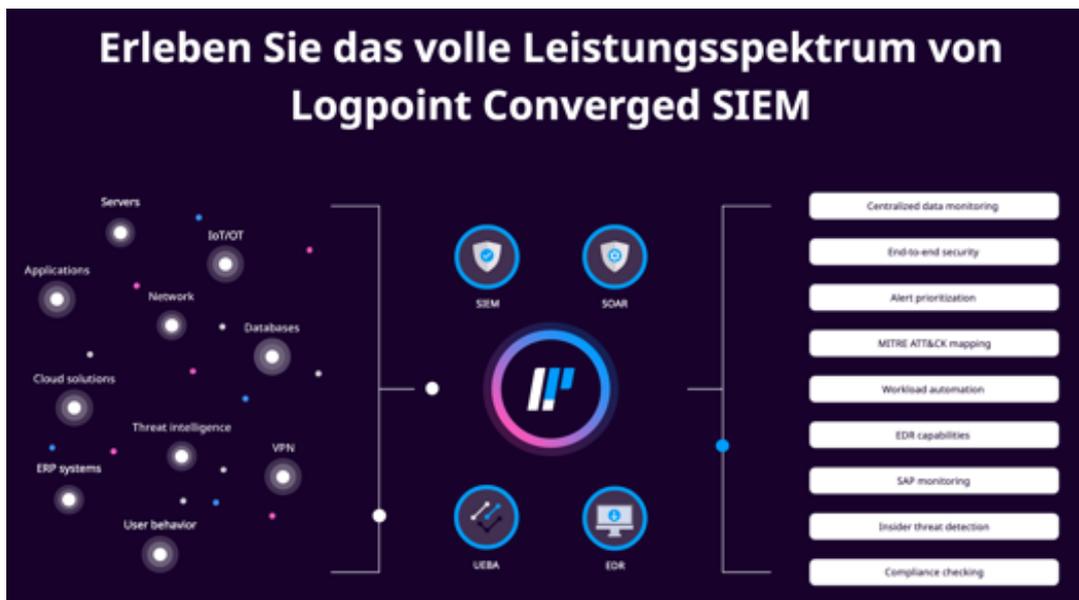


Abbildung 38: Logpoint SIEM

Centralized Data Monitoring wird das Benutzerverhalten analysiert, um Anomalien und potenzielle Bedrohungen zu erkennen, und die Server werden kontinuierlich auf verdächtige Aktivitäten und Sicherheitsverstöße überwacht.

End-to-End Security bietet umfassenden Schutz auf jeder Ebene Ihrer IT-Infrastruktur, von Endgeräten bis hin zu Servern.

Alert Prioritization ermöglicht Alarmpriorisierung durch Auswertung und Priorisierung von Sicherheitswarnungen sowie eine schnelle Reaktion auf kritische Ereignisse.

MITRE ATT&CK-Mapping ist eine Adaption des MITRE ATT&CK-Framework zur Identifizierung und Kategorisierung von Angriffstechniken.

Workload Automation automatisiert Sicherheitsaktivitäten, um manuelle Aufgaben zu entlasten und die Reaktion zu beschleunigen.

Endpoint Detection and Response (EDR) Capabilities erweitern die Fähigkeit, Sicherheitsereignisse auf Endpunktgeräten zu erkennen und darauf zu reagieren.

SAP Monitoring Server ist auf die Überwachung von SAP-Anwendungen zur Erkennung von Sicherheitsbedrohungen spezialisiert.

Insider threat detection funktioniert durch die Identifizierung von Bedrohungen, die möglicherweise von internen Akteuren ausgehen.

Compliance Checking versteht man Kontrollen zur Einhaltung von Sicherheitsstandards und Compliance-Richtlinien.³⁹

vgl.³⁹ (Logpoint)

2.3.3.2. Logpoint mit KI

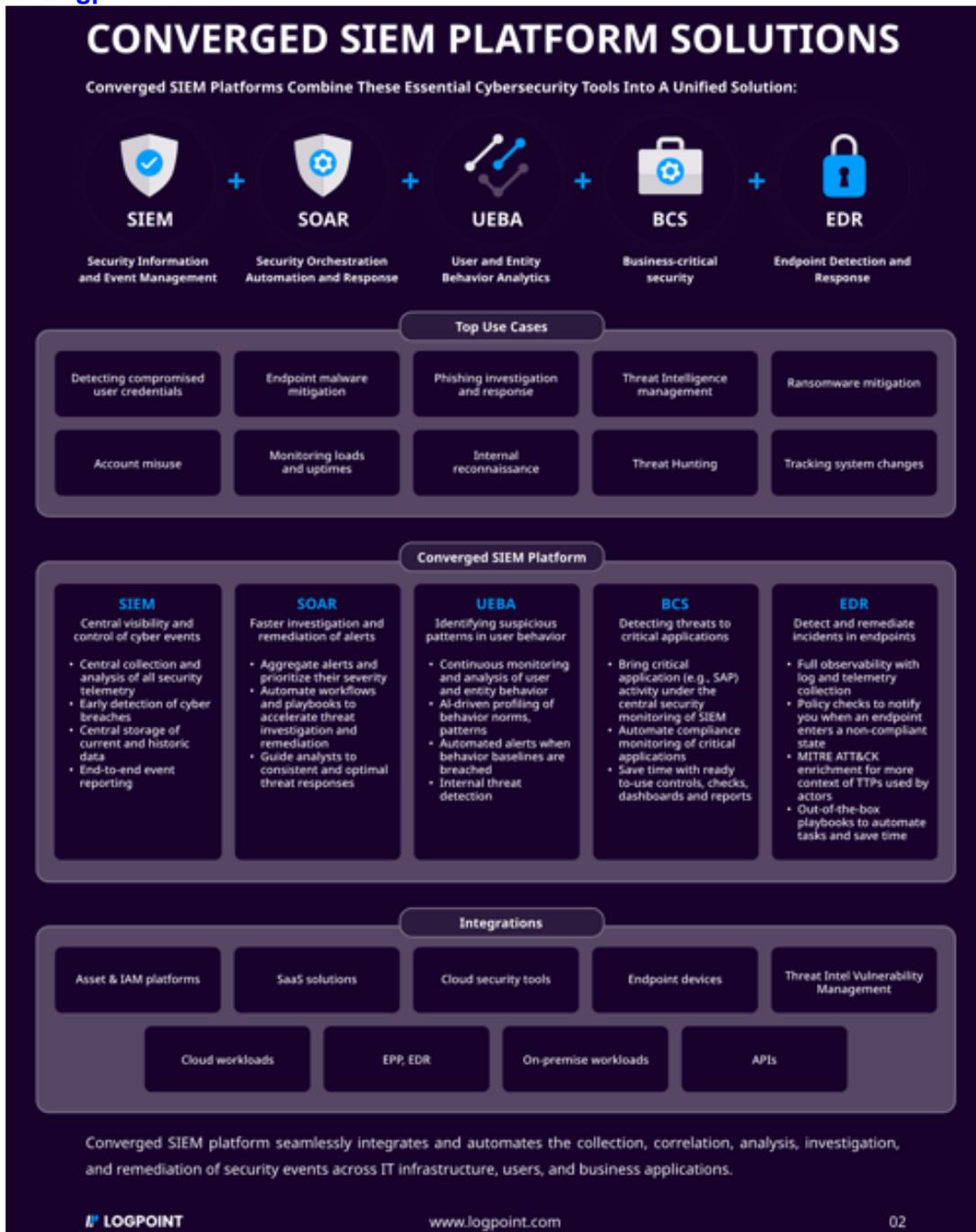


Abbildung 39: Logpoint SIEM Platform Solution

SIEM-, SOAR- und UEBA-Systeme besitzen unterschiedliche Fähigkeiten in den Bereich der Cybersicherheitsoperationen und tragen zur Cybersicherheitslage eines Unternehmens bei. Viele Unternehmen erstellen als erstes eine SIEM-Plattform, dann SIEM und UEBA für detailliertere Bedrohungsintelligenz und fügen eventuell dann SOAR hinzu, um die Reaktion auf Bedrohungen zu automatisieren.

Standalone-Produkte erfordern bei der Integration Zeit und Geld, um Best-in-Class-Lösungen mit eigenständigen Systemen zu integrieren, erhöht Komplexität und Kosten. **Business Critical Security (BCS)** wurde entwickelt für Cyberangriffen und richten sich zunehmend gegen kritische Unternehmensanwendungen mit angereicherten und wertvollen Daten, als nur auf traditionelle IT-Infrastruktur. Diese Anwendungen sind für den reibungslosen Geschäftsbetrieb und Kundenbetreuung unverzichtbar. Moderne konvergierte SIEM-Plattformen beinhalten solche geschäftskritischen Anwendungen, wie z.B. SAP mit ihrer Sicherheitsüberwachung und können dadurch komplexe Angriffe aufdecken, die in diesen Anwendungen beginnen oder stattfinden, und die sonst schwer erkennbar wären.

Endpoint Detection and Response (EDR) werden häufige Angriffe werden an den Endpunkten mit Schwachstellen durchgeführt, um Bedrohungen auf diesen Geräten zu erkennen und schnell zu reagieren. Logpoint ermöglicht mit dem AgentX Tool, einem nativen Endpunkt-Agenten und den erweiterten EDR-Funktionen, die Sicherheit des gesamten Unternehmens zu gewährleisten. In Verbindung mit SIEM und SOAR garantiert AgentX eine automatisierte Erkennung, Untersuchung und Reaktion auf Endpunktvorfälle, garantiert somit die Einhaltung von Compliance-Vorgaben und erstellt Protokolle von allen Geräten. Das Ereignis wird angereichert mit relevanten Kontext- und Betriebsdaten, sodass das SOC-Teams im grossen Umfang Analysen erhalten, die darauf abzielen, effizientere Lösung von Vorfällen zu erstellen.

In den integrierten **SIEM-Plattform sind SIEM-, SOAR-, UEBA-, BCS-Tools und EDR-** Kompetenzen bereits integriert und bereit, abgestimmt zu arbeiten und ihre Flexibilität bleibt dabei bestehen, um die Tools auf einmal oder schrittweise zu implementieren.⁴⁰

Diese Tabelle zeigt einige wichtige Anwendungsfälle für SIEM, SOAR, UEBA, BCS und EDR-Tools auf.⁴¹

Use case	SIEM	SOAR	UEBA	BCS	EDR
Insider Threat Detection	Advanced analytics help you find malicious insiders that are difficult to uncover with traditional detection methods. Prevent security incidents before they cause irrevocable damage.				
Fraud and Risk Monitoring	Continuously monitor the behaviors of users, computers, and IoT devices and link events across your environment to uncover anomalies. Automated playbooks trigger the necessary action plan.				
Anomalous File Sharing and Data Exfiltration	Detect anomalous user activity based on historical and expected work patterns. Correlate data from users and other sources to identify lateral movement and prevent data from leaving the system.				
Malware Detection and Mitigation	Reconstruct the events that led to malware infection using the data at your disposal. Document the full scope of the breach, prevent additional systems from becoming infected and initiate playbooks to remediate incidents.				
Internal Reconnaissance	Gather evidence from network infrastructure to uncover anomalous behavior. When you investigate an alert, contextual information from related incidents and threat intel is automatically added to help detect and prevent attackers from gaining information.				
Regulatory Compliance Monitoring	Automated risk assessment monitoring centralizes and analyzes data across the organization to meet compliance standards. Comprehensive reporting proves adherence to compliance and regulatory frameworks.				
Intellectual Property Theft	Automatically investigate incidents to determine causation. Case management identifies relevant data to help you uncover where the attacker infiltrated the system and what was accessed.				
Incident Prioritization and Triage	Quickly investigate incidents with root cause analysis, risk level priority, and automatic event context. Find and analyze relevant data to uncover threats or suspicious activity.				
Compromised Account and Misuse Detection	Prevent unauthorized account use by anyone other than the account holder. Monitor how employees behave within the system and detect any unauthorized account use by account holders.				
Threat Hunting	Threat hunting capabilities provide in-depth hunting and analysis through enrichment, correlation, machine learning and threat intelligence. Search proactively for cyberthreats that might otherwise evade detection.				
Phishing Investigation and Response	Track who received phishing emails, clicked on any malicious links, or replied to them and take immediate action to minimize damage. Automatic playbooks and machine learning help minimize impact on your organization.				
Vulnerability management	Risk-based prioritization based on anomalous user behavior, threat intel and machine learning, helps detect vulnerabilities in your organization. Automated playbooks help analyze and remediate each vulnerability.				

Abbildung 40: wichtige Anwendungsfälle

vgl.⁴⁰ (LogPoint, 2023)

vgl.⁴¹ (LogPoint, 2023)

User and Entity Behavior Analytics (UEBA) fördert die Erkennung von Bedrohungen, um das Verhalten von Benutzern und Entitäten in einem Netzwerk zu analysieren. Dabei wird maschinelles Lernen verwendet, um normale Verhaltensmuster zu erzeugen und Abweichungen davon zu analysieren, um die möglichen Sicherheitsrisiken anzuzeigen. Dies steigert die Bedeutung der Integration von UEBA in bestehende Sicherheitsinfrastrukturen, insbesondere in Verbindung mit SIEM (Security Information and Event Management).

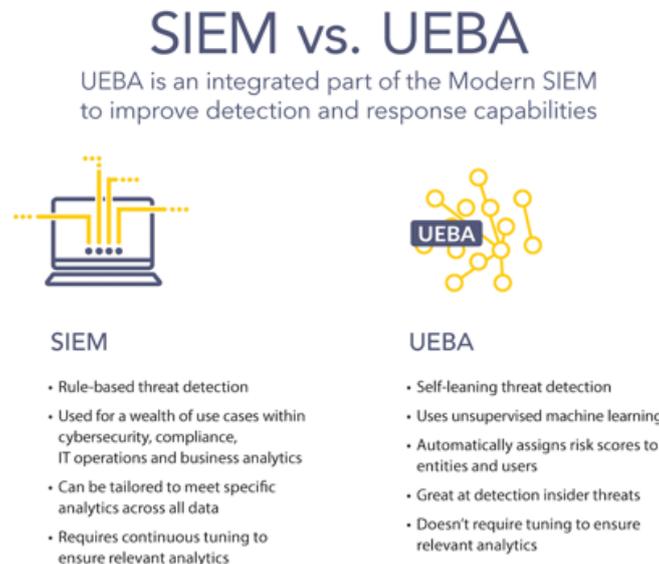


Abbildung 41: SIEM vs. UEBA

UEBA hilft zur Erkennung von Insider-Bedrohungen und externen Angriffen, die oft auf kompromittierten Benutzerkonten hinzielen. Durch eine kombinierte Nutzung von UEBA und SIEM für eine umfassende Sicherheitsstrategie ein grosser Vorteil.⁴²

Vorteile von UEBA sind **automatisierte Bedrohungserkennung** durch ML und daraus können Verhaltensanalysen im Unternehmen schnell eingesetzt werden, um kompromittierte Konten, Brute-Force-Angriffe und Verletzungen der Datensicherheit zu erfassen und somit begrenzte personelle Ressourcen effektiv anderswertig einzusetzen; Risikominderung mit vorausschauende Erkennung von kompromittierter Anmeldeinformationen erlauben es Unternehmen, Risiken und potenzielle Datenverluste zu reduzieren, indem der interne Zugriff auf das Netzwerk durch Cyberkriminelle unterbunden wird; **kürzere Reaktionszeiten** im UEBA senkt die Reaktionszeit auf Angriffe mit Hilfe einer akkurate Risikobewertung, sodass das Sicherheitsteams Einbruchversuche schnell erkennt und darauf reagiert, um Schäden im Unternehmen einzugrenzen sowie die Verhaltensanalysen unterstützten bei der **Vermeidung und Reduzierung von Fehlalarmen**, sodass das Security-Teams sich auf Aktivitäten mit dem größten Risiko zu konzentrieren und so die Effizienz ihrer Reaktion auf kritische Bedrohungen zu verbessern.

UEBA-Grenzen und Lösungen bei diesem Cybersecurity-Tools ermöglichen es als integrierte Plattform einen extensiven Schutz und tiefgreifendere Umsetzung, in Unternehmen, eine Auswahl auf Qualität der Daten, Integration mit anderen Sicherheitslösungen und Anpassungsfähigkeit zu achten, um effektiv vor Cyberbedrohungen geschützt zu sein. UEBA erkennt gut Insider-Bedrohungen, aber es kann Schwierigkeiten bei der Erkennung spezifischer Angriffstypen wie Bild-Malware geben, was zusätzliche spezialisierte Tools erfordert.

vgl.⁴² (CORTEX)

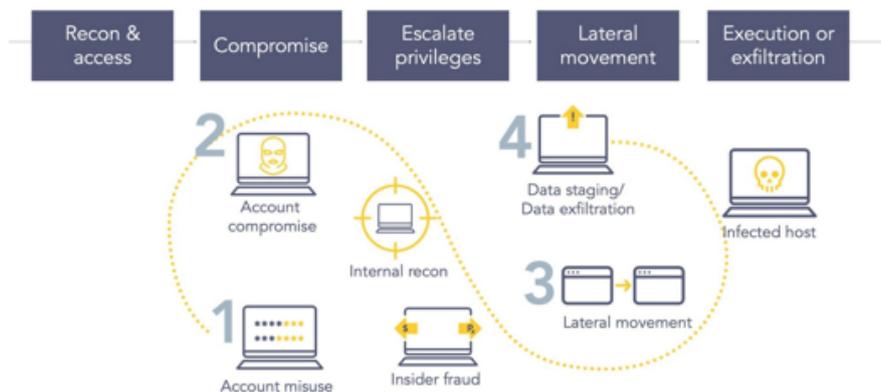


Abbildung 42: UEB- Erkennung

Logpoint SIEM verwendet UEBA und bietet somit integrierte Lösungen mit umfassender Datenmustererkennung an, benötigt weniger Kalibrierungseinstellungen und -anpassungen sowie erstellt transparente Kostenstrukturen. Logpoint erfasst Daten in einer gemeinsamen Sprache für maximale Effizienz von ML und automatisiert diese. Mit sofortiger Einsatzbereitschaft und präziser Bedrohungserkennung ist dieses SIEMs-Tool eine kosteneffektive und intuitive Lösung für Unternehmen, da fortschrittliche Bewertungsmethoden es ermöglichen, Zeit und Ressourcen auf die größten Risiken zu konzentrieren, ohne versteckte Kosten entstehen zu lassen.⁴³

Best Practices for UEBA – Considerations for success



Define Use Cases:

- Clearly define the use cases
- Describe the desired output from the UEBA for each of these cases
- GOAL: Ensure UEBA delivers value against crucial pain points



Consider data sources:

- Make sure UEBA can support your priority use cases' necessary data sources
- Check if these use cases can be implemented out-of-the-box with prepackaged analytics
- Determine the expected effort level for custom use cases
- GOAL: Ensuring required coverage for anomaly detection across the entire organization



Proof of Concept:

- A UEBA POC can last 30 days to construct baselines using live or historical data.
- Evaluate and review anomalies detected by UEBA and their legitimacy
- GOAL: proofing capabilities between between technology and customer requirements

Abbildung 43: Best Practices for UEBA

2.3.3.3. Use Case Logpoint

Die immer größer und komplexer werdende Cyber-Bedrohungslandschaft führt häufig dazu, dass sich Malware-Akteure anpassen und neu positionieren müssen. Ein aktuelles Beispiel hierfür ist **DarkGate**, das als Folge des Zusammenbruchs der Qbot-Plattform erstellt wurde. Darkgate ist eine komplexe, vielseitig einsetzbare Malware, die den Diebstahl vertraulicher Informationen sowie die Verwendung von Kryptowährung zum Mining dieser Informationen ermöglicht. Die Malware wird von einem Entwickler mit dem Pseudonym „RastaFarEye“ veröffentlicht und über Phishing-Kampagnen, Malvertising und SEO-Poisoning (Search Engine Optimization Poisoning) verbreitet. Besonders zum Einsatz kommt die Verwendung von Autolt, einer Softwareanwendung für die Windows-Plattform, die zur Automatisierung sich wiederholender Aufgaben verwendet wird. Phishing bleibt nach wie vor die primäre Infektionsmethode. Täter wenden verschiedene Taktiken wie gefälschte Browser-Updates und veränderte Microsoft Teams-Einladungen an. Andere Verbreitungsmethoden wie

vgl.⁴³ (LogPoint, 2020)

Malvertising und SEO-Purging nutzen normalerweise Tools wie den Advanced IP Scanner. Die DarkGate-Malware wird über verschiedene Verbreitungsmethoden an Kriminelle vermarktet und von verschiedenen Personen verwendet. Ein häufiges Ziel von Angriffen ist die Infektion von Computern mit der Absicht, in IT-Systeme einzudringen, wie der Einsatz von Scannern bei diesen Angriffen zeigt. Diese DarkGate-Infektionskette verwendet verschiedene Dateiformate wie **.msi**, **.lnk** und **.vbs**, um ersten Zugriff auf das beabsichtigte System zu erhalten. Anschließend werden die sogenannten „**Living Off The Land**“-Binärdateien (LOLBins) und zusätzliche Malware verwendet, um die Hauptkomponente der Malware zu installieren.⁴⁴

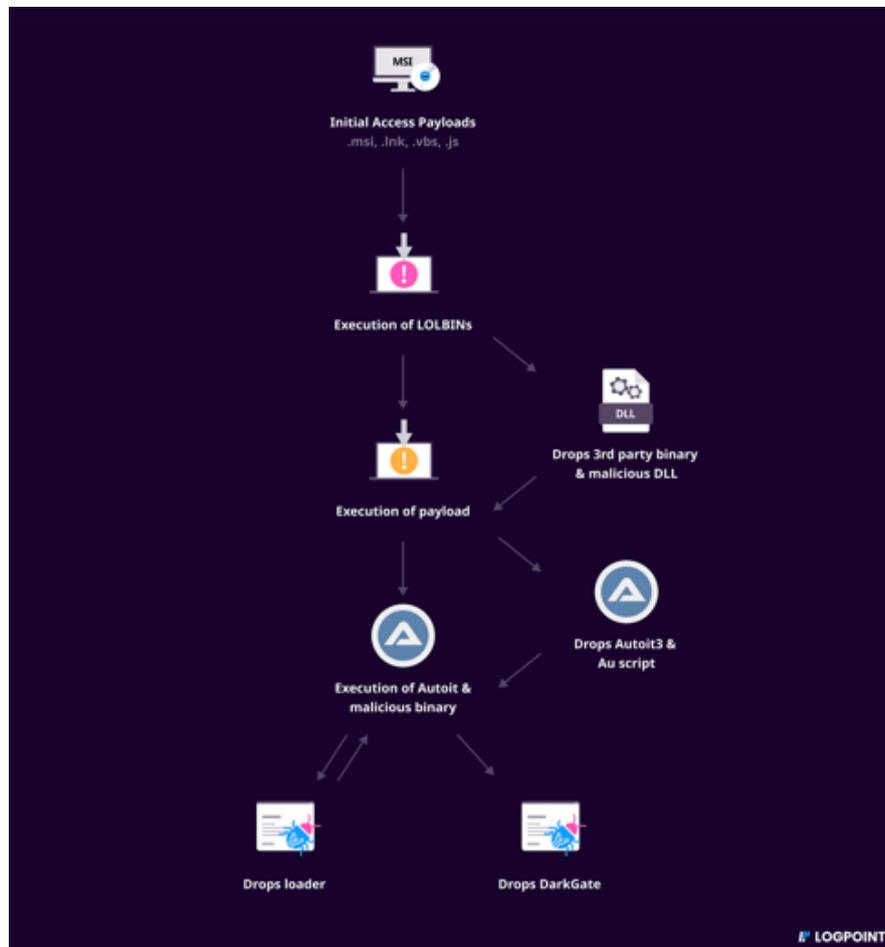


Abbildung 44: DarkGate Infektionskette

Darkgate stellte zunächst verschiedene Arten von **Payloads** bereit, die den Zugriff ermöglichen, darunter **.msi**, **.lnk**, **.vbs**, **.js** und so weiter. Sobald diese in den Besitz gelangt sind, nutzen die Opfer sie, um die Kriminellen zu töten. Die Standard-Apps sind mit den zugehörigen Dateitypen verbunden. Darkgate benutzt die Living Off The Land Binaries (LOLBins), um die Payload um eine weitere Komponente zu erweitern. Dies wird häufig als Dateien im Archivformat (**.cab**) dargestellt, aus denen zusätzliche Payload-Stufen extrahiert werden können. Diese bestehen aus externen Dateien und ihren zugehörigen böswärtigen Kopien, die das Laden von DLLs erleichtern. Alternativ können Payloads, die abgeworfen werden, zur dritten Stufe hinzugefügt werden. Wenn die Ausführung erfolgt, wird der Download von ausführbaren Dateien ausgelöst. Nach dem Laden der DLLs lädt DarkGate die Autoit3-Binärdateien und -Skripte und stellt dann erneut eine Verbindung zum Kontrollzentrum her, um die zugehörigen Dateien zur Verfügung zu stellen. Letztendlich veranlasst DarkGate **Autoit3.exe** dazu, schädliche AU-Skripte auszuführen, die Anweisungen enthalten, die zur Extraktion und Platzierung des Loaders führen, bevor das primäre Modul aktiviert wird. DarkGate und sein Loader führen Sicherheitsbewertungen durch, die das Vorhandensein bestimmter Ordner und Prozesse identifizieren, bevor schädliche Funktionen ausgeführt werden. Zur Reaktion auf DarkGate-Malware stehen mehrere Methoden zur Verfügung,

vgl.⁴⁴ (Logpoint)

darunter das Entfernen von Dateien, wobei DarkGate-bezogene Dateien mithilfe eines automatisierten Prozesses analysiert und gelöscht werden; der Beendigungsprozess, der automatisch mit aktiven Malware-Prozessen endet; die Registrierung, die automatisch gelöscht wird; und die Host-Isolierung, die eingesetzt wird, um infizierte Hosts zu isolieren und so die Verbreitung von Malware zu reduzieren. Diese Maßnahmen sind automatisiert und ermöglichen eine wirksame Reaktion auf DarkGate-Infektionen. Zur Verteidigung gegen Darkgate und andere ähnliche Gefahren wird empfohlen, dass Mitarbeiter darin geschult werden, Social-Engineering-Angriffe zu erkennen, eine mehrstufige Authentifizierung und häufige Kennwortänderungen zu verwenden, das Prinzip der geringsten Privilegien für Benutzer umzusetzen, Schutzsysteme zu verwenden, die auf Software und Geräten basieren, und einen Vorfallreaktionsplan für eine sofortige Reaktion bereitzuhalten. Für ein zentralisiertes Protokollierungs- und Überwachungssystem bietet Logpoint eine konvergente SIEM-Plattform. Nachdem die Infektion des Hosts offiziell erkannt wurde, können Analysten das Protokoll „**Isolate-Unisolate Host**“ verwenden, um den Host zu isolieren und verhindert das Potenzial für weitere Bewegungen und Pivots.⁴⁵

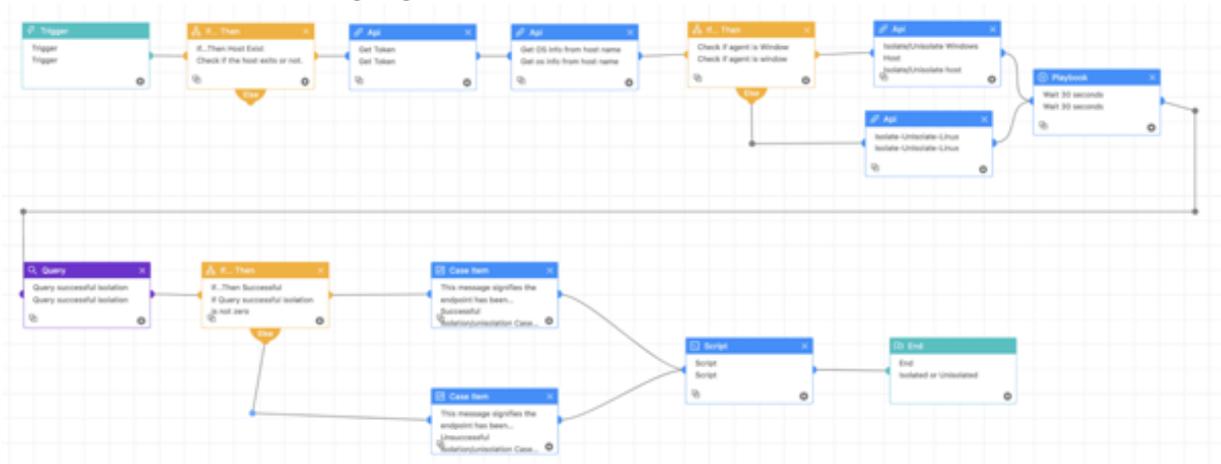


Abbildung 45: Logpoint AgentX Isolate-Unisolate Host

3.3.3.4. Demo von Logpoint

Logpoint Test-Demo unter [Book a demo](#) zu buchen:⁴⁶



Abbildung 46: Logpoint SIEM Demo

vgl.⁴⁵ (Logpoint, et al., 2024)
vgl.⁴⁶ (Logpoint)

2.3.4. LogRhythm

Im folgenden Abschnitt 3.2.4.3. wird das Tool LogRhythm etwas genauer beschrieben:⁴⁷

LogRhythm ist eine SIEM-Lösung von LogRhythm ist die LogRhythm SIEM Plattform, die mehrere Add-on-Komponenten umfasst, um Endpoint-, Netzwerk- und Benutzerverhaltensanalysen bereitzustellen. Die Mehrheit der SIEM-Kunden von LogRhythm befindet sich in Nordamerika und Europa, während der Rest in der Region Asien/Pazifik, im Nahen Osten, in Afrika und Lateinamerika ansässig ist. Die Kundenbasis neigt zu mittelständischen Unternehmen und kleineren Organisationen, obwohl auch große Unternehmen die LogRhythm SIEM erworben haben. Es gibt eine Cloud-Option, aber die meisten Kunden haben ihre SIEM-Lösung lokal implementiert. Die Lizenzierung erfolgt entweder auf Basis einer unbefristeten Lizenz (nach durchschnittlicher Anzahl von Nachrichten pro Sekunde und Tag) oder auf Abonnementbasis (nach Anzahl der Mitarbeiter).

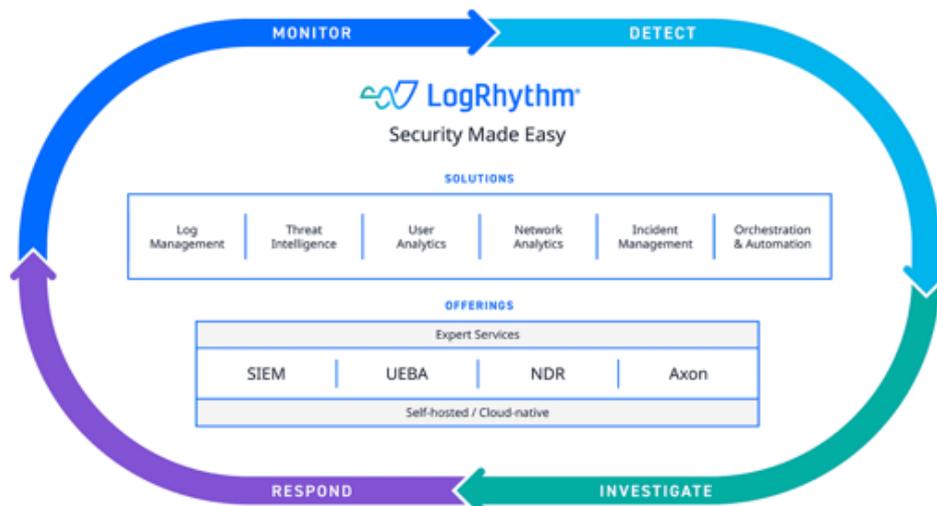


Abbildung 47: LogRhythm

2.3.4.1. wichtige Funktionen von LogRhythm

Sicherheitsoperationen werden optimiert über den Bedrohungslebenszyklus und reduziert die **Mean Time To Detect (MTTD)** und die **Mean-Time-To- Reaktion (MTTR)**.

Zentrale Fallverwaltung ermöglicht es Sicherheitsteams, ihre Effizienz zu steigern und schneller auf Vorfälle zu reagieren.

Vollautomatisierte Playbooks automatisieren eine Vielzahl von Aufgaben, während halb-automatische, genehmigungsbasierte Reaktionsaktionen eine Genehmigungsbasis für bestimmte Reaktionen bieten, um sicherzustellen, dass angemessene Maßnahmen ergriffen werden.

SmartResponse automatisiert nicht nur Aufgaben, sondern bietet auch breitere Handlungsmöglichkeiten, um den spezifischen Anforderungen und Bedrohungen einer Organisation gerecht zu werden und damit unterstützt LogRhythm Unternehmen dabei, ihre Sicherheitsreife zu steigern und proaktiv auf aktuelle Bedrohungen zu reagieren.

Log- und Ereigniserfassung kann Protokolle und Ereignisse aus verschiedenen Quellen sammeln, darunter Netzwerke, Betriebssysteme, Anwendungen und Sicherheitsgeräte. Es ermöglicht die zentrale Erfassung und Speicherung von Protokolldaten.

Echtzeit-Überwachung wird durchgeführt, um verdächtige Aktivitäten oder Sicherheitsvorfälle sofort zu identifizieren. Das umfasst die Überwachung von Netzwerkverkehr, Benutzeraktivitäten und Systemprotokollen.

Bedrohungserkennung nutzt fortschrittliche Analysetechniken, um Anomalien und verdächtiges Verhalten zu erkennen. Das System kann potenziell schädliche Aktivitäten oder Sicherheitsverletzungen frühzeitig identifizieren.

vgl.⁴⁷ (LogRhythm)

Automatisierte Reaktion ermöglicht es, auf erkannte Bedrohungen automatisierte Reaktionen einzurichten. Dies kann von der Sperrung von Benutzerkonten bis zur Isolierung von betroffenen Systemen reichen.

Incident Response unterstützt den gesamten Incident-Response-Prozess. Dies beinhaltet die Erfassung von Informationen über Sicherheitsvorfälle, die Analyse von Ursachen und Auswirkungen sowie die Bereitstellung von Werkzeugen für die effiziente Reaktion und Gegenmaßnahmen.

Compliance-Management hilft bei der Einhaltung von Sicherheitsrichtlinien und -standards durch die Überwachung und Dokumentation von Sicherheitsereignissen. Dies ist besonders wichtig für Organisationen, die branchenspezifische Vorschriften einhalten müssen.

Benutzer- und Zugriffsüberwachung bietet Funktionen zur Überwachung von Benutzeraktivitäten und Zugriffen, um unautorisierte oder verdächtige Handlungen zu erkennen.

Reporting und Analyse ermöglicht die Erstellung von Berichten und Analysen basierend auf den gesammelten Protokoll Daten. Dies unterstützt Sicherheitsanalysten und Führungskräfte bei der Bewertung der Sicherheitslage.

Integrierte Threat Intelligence integriert Informationen aus Threat-Intelligence-Quellen, um aktuelle Bedrohungsdaten in die Analysen einzubeziehen und so die Erkennung von fortgeschrittenen Bedrohungen zu verbessern.

Skalierbarkeit und Flexibilität sind darauf ausgelegt, mit dem Wachstum der Organisation zu skalieren und sich an verschiedene Umgebungen anzupassen.

Umfangreiche Reseller verfügen über ein starkes Team von Vertriebspartnern in jeder wichtigen Region weltweit und dies unterstützt Managed Service Providern unterstützt, um ressourcenbeschränkten Käufern bei der Verwaltung und Überwachung ihres SIEM zu helfen.

Optionsmöglichkeiten für Pilotprojekte und Proof of Concept (PoC) können Käufer von verschiedenen Arten von Pilot- und PoC-Programmen profitieren, angefangen von Vorbereitungs-Workshops über gehostete, szenariobasierte Testfahrten bis hin zu "Versuchen und Kaufen"-Optionen.

Untersuchungs- und Fallmanagement-Workflow bietet ausgereifte und verfeinerte Funktionen für Untersuchung und Fallmanagement, die Kontext zusammenstellen und Benutzern ermöglichen, eine Grundlage für die Fallentscheidung zu schaffen.

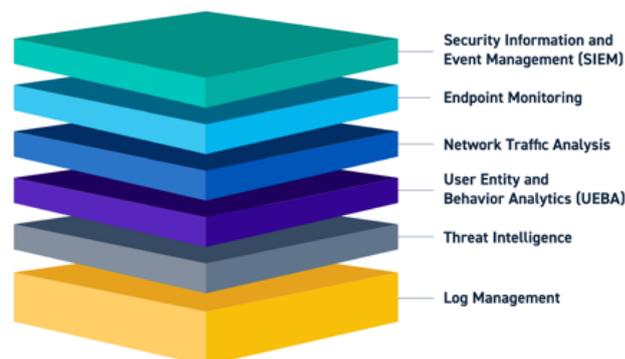


Abbildung 48: Evolution of SIEM-Software

2.3.4.2. LogRhythm mit KI

LogRhythm ist ein Anbieter von Sicherheitsinformations- und Ereignismanagementlösungen (SIEM), die Unternehmen unterstützend bei der Erkennung von Bedrohungen hilft, untersucht und darauf reagiert. LogRhythm verwendet Künstliche Intelligenz (KI), um Sicherheitsanalysen zu verbessern und die Reaktionszeit auf potenzielle Bedrohungen zu verkürzen.

Verhaltensanalyse nutzt KI, um das normale Verhalten von Benutzern, Systemen und Anwendungen zu analysieren und Anomalien zu erkennen und zu bewerten, die auf potenzielle Sicherheitsbedrohungen hinweisen könnten. Durch die kontinuierliche Überwachung und Analyse von Datenströmen werden verdächtige Aktivitäten in Echtzeit ermittelt.

Automatisierte Bedrohungserkennung werden mit der Unterstützung von KI und ML potenzielle schädliche Aktivitäten automatisch erkannt und Prioritäten gesetzt und somit

können Sicherheitsanalysten sich auf die dringendsten Bedrohungen konzentrieren, um Erkennungs- und Reaktionszeiten zu verkürzen.

Erweiterte Analysen und Forensik dienen mit dem Einsatz von KI bei Analyseprozessen werden tiefgreifende Untersuchungen von Sicherheitsvorfällen durchgeführt und durch eine Kombination von Verhaltensanalysen mit ML erkannt, die möglicherweise von herkömmlichen Analysemethoden übersehen werden.

Bedrohungsintelligenz werden externe Bedrohungsdaten und -informationen in dieser Plattform mit KI genutzt, um Informationen zu analysieren und Bedrohungen proaktiv zu herauszufiltern, um sich vor bekannten und aufkommenden Bedrohungen zu schützen.

Automatisierte Reaktionen bietet die Möglichkeit durch die Kombination von Sicherheitsanalysen mit KI automatisierte Reaktionen auf bestimmte Sicherheitsvorfälle einzurichten. Dies kann von automatisierten Benachrichtigungen bis hin zur Ausführung von Maßnahmen zur Eindämmung oder Abwehr von Bedrohungen reichen.⁴⁸

Die Integration von KI und ML in die SIEM-Lösungen von LogRhythm bietet die Effizienz der Sicherheitsoperationen zu verbessern, die Erkennung von Bedrohungen zu stärken und die Reaktionszeit auf Sicherheitsvorfälle effektiv zu verkürzen.⁴⁹

Weitere Use Cases sind laut DevOpsSchool für Log Rhythm sind:⁵⁰

Threat Detection and Alerting überwacht fortlaufend Netzwerk- und Systemaktivitäten in Echtzeit, analysiert Protokolle und Ereignisse, um verdächtiges oder böses Verhalten zu erkennen und generiert Warnungen und Benachrichtigungen bei potenziellen Bedrohungen.

Incident Investigation können Sicherheitsteams verwendet werden, um Sicherheitsvorfälle zu untersuchen, durch das Analysieren von historischen Protokoll- und Ereignisdaten, um die Ursachen von Vorfällen zu agnoszieren und deren Auswirkungen zu verstehen.

Security Event Correlation mit Korrelations-Engine werden Daten aus verschiedenen Quellen verbunden, um komplexe Angriffsmuster und fortschrittliche Bedrohungen zu entdecken und diagnostizieren von Bedrohungen, die einzelnen Sicherheitskontrollen entfallen können.

Compliance Management unterstützt Unternehmen bei der Einhaltung regulatorischer Anforderungen mit vordefinierten Compliance-Berichten und automatisierter Compliance-Überwachung.

User and Entity Behavior Analytics (UEBA) analysiert das Verhalten von Benutzern und Entitäten, sodass Anomalien und potenziell böse Aktivitäten erkannt werden, was besonders wichtig ist für die Erkennung von Insider-Bedrohungen und verschleierte Konten.

Vulnerability Management (VMS) ist ein Schwachstellenbewertungstool, um Schwachstellen basierend auf dem potenziellen Einfluss auf die Sicherheit der Unternehmen zu priorisieren und zu beheben.

Advanced Analytics verwendet erweiterte Analysefunktionen mit ML und Verhaltensanalytik, um Bedrohungen und Anomalien zu erkennen, die herkömmlichen Erkennungsmethoden entfallen können.

Incident Response Orchestration unterstützt die Vorfalleinstufung und -orchestrierung durch Workflow-Automatisierungstools, um den Reaktionsprozess zu verbessern und sicherzustellen, dass zeitnahe Maßnahmen ergriffen werden.

Log Management and Analysis ist eine zentrale Plattform für die Protokollverwaltung, die Protokolle und Ereignisse von verschiedenen Quellen sammelt, speichert und analysiert in Netzwerkgeräten, Servern und Anwendungen.

Cloud Security Monitoring dient der Sicherheitsüberwachung auf Cloud-Umgebungen auszuweiten, um mit Cloud-Plattformen und -Diensten zu integrieren, um Protokoll- und Ereignisdaten zu analysieren.

vgl.⁴⁸ (logrhythm)

vgl.⁴⁹ (logrhythm)

vgl.⁵⁰ (Ashwani, 2023)

Insider Threat Detection hilft bei der Erkennung interner Bedrohungen, um Benutzeraktivitäten und das Verhalten zu überwachen und so ungewöhnliche oder verdächtige Handlungen zu erkennen.

IoT Security / Internet der Dinge (IoT) werden IoT-Geräte und -Netzwerke überwacht und gesichert, um ihre Aktivitäten und ihr Verhalten zu analysieren.

Integration von Endpunkterkennung und -antwort (EDR) wird die Endpunktsicherheit, von der Endpunkt-Daten mit Netzwerk- und Systemereignissen korreliert werden, verbessert.

Network Traffic Analysis (NTA) dient der Analyse von Netzwerkverkehr, um Netzwerkaktivitäten zu überwachen und zu analysieren, unbefugten Zugriff zu identifizieren und Anomalien zu erkennen.

LogRhythm ist eine **umfassende Sicherheitsplattform**, die Daten aus verschiedenen Quellen wie Netzwerkgeräten, Servern, Anwendungen und Cloud-Services sammelt, analysiert und überwacht. Die gesammelten Daten werden normalisiert und in Echtzeit analysiert, um Sicherheitsbedrohungen und Anomalien zu erkennen und bei Vorfällen Sicherheitsereignisse zu erstellen und zu blockieren.

Die **LogRhythm-Architektur** besteht aus verschiedenen Komponenten und arbeitet zusammen, wie Datensammler, Datenverarbeitung, Korrelations-Engine, Alarmierung und Benachrichtigung, Benutzeroberfläche, Datenspeicherung, Referenzdaten und externe Integration, um um Protokoll- und Ereignisdaten aus verschiedenen Quellen zu sammeln, zu analysieren, zu normalisieren und Anomalien zu überprüfen. LogRhythm verfügt über eine benutzerfreundliche Oberfläche für Sicherheitsanalysten, um Vorfälle zu untersuchen, Berichte zu erstellen und Alarme zu generieren. Die Architektur ist skalierbar und hochverfügbar, um eine kontinuierliche Überwachung von Sicherheitsbedrohungen sicherzustellen und eine umfassende Plattform für Sicherheitsoperationen und Vorfälle zu bieten.⁵¹

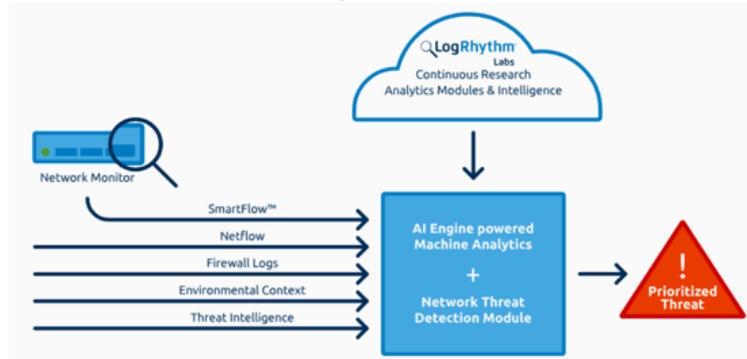


Abbildung 49: LogRhythm Architektur

2.3.4.3. Demo von LogRhythm

LogRhythm Test-Demo unter [Book a demo](#) zu buchen.⁵²



Abbildung 50: LogRhythm SIEM Demo

vgl.⁵¹ (Ashwani, 2023)

⁵² (LogRhythm)

2.3.5. SolarWinds

Im folgenden Abschnitt 3.2.4.4. wird das Tool SolarWinds etwas genauer beschrieben.⁵³

SolarWinds legt Fokus auf Produktsätze für Application Performance Monitoring (APM) und Observability sowie die Bereitstellung von Diensten über SaaS und Überwachung vor Ort. SolarWinds beinhaltet geplante Verbesserungen für seine Observability Suite, AIOps-Optimierungen und für die Einführung eines Moduls für Cloud-Kostenanalysten.

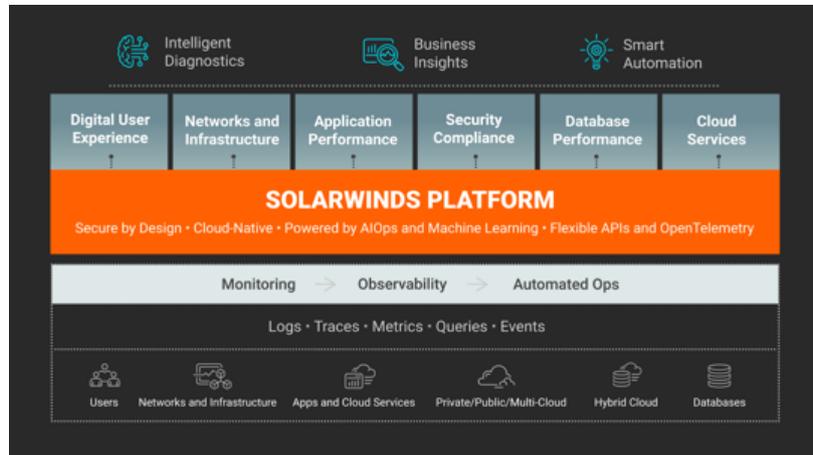


Abbildung 51: SolarWinds SIEM

2.3.5.1. wichtige Funktionen von SolarWinds mit KI

Portfolio bietet eine umfangreiche Produktsuite, die weitreichende IT-Betriebsbelastungen abdeckt und neben APM für die Überwachung von Datenbanken, Netzwerken und Sicherheitstools beinhaltet.

Problembehandlung und Diagnose bei auftretenden Problemen ermöglicht die Schnittstelle Benutzern die Diagnose von Netzwerk- oder Systemstörungen, um schnell Maßnahmen zur Fehlerbehebung erstellen zu können.

Preisgestaltung für KMU wird von kleinen und mittelständischen Unternehmen attraktiv angenommen, die besonders mit begrenzter Erfahrung in der Implementierung von APM-Tools haben. Die neu gestaltete Preisstruktur für Hybrid Cloud Observability soll Kunden helfen, Ausgaben für Observability-Lösungen besser zu verstehen und zu kontrollieren, und besonders für Unternehmen mit knappem Budget eine attraktive Option.

SaaS-Lösung als die neue SolarWinds Observability Suite bietet ein SaaS-Bereitstellungsmodell für seine Observability-Lösung an und führt dazu bei, dass es zu verkürzten Implementierungszeiten und einem schnelleren Return on Investment für neue Implementierungen erstellt wird.

Priorität für Observability wird verstärkt auf KI-gesteuerte Observability-Lösungen eingesetzt, um Verluste durch Ausfälle zu minimieren und proaktiv auf Probleme in Apps und Infrastrukturen zu reagieren.

Fokus auf Datenbankprobleme wird das Team KI und Automation nutzen, um die Gesundheit, Stabilität und Skalierbarkeit von über 300 Datenbanken zu gewährleisten und Ausfälle zu vermeiden.

Einführung von AIOps eingeführt für den IT-Betrieb (AIOps) mit KI, um in komplexen IT-Umgebungen Daten zu integrieren, Leistung zu optimieren und IT-Teams zu entlasten.

Beschleunigung von IT-Service-Management (ITSM) werden mit KI-gesteuerten Tools beschleunigt, dabei werden die Ausfallzeiten um 21% reduziert und die Bearbeitungszeit von Zwischenfällen um 23% verkürzt wird.⁵⁴

Hierbei werden Funktionen, wie Incident Management, Service Requests, Automation, Workflows, Runbooks, Problem-Management, Change-Management, Asset-Management, Procurement und ein raffiniertes Configuration Management Database (CMDB) verwendet.

vgl.⁵³ (SolarWinds)

vgl.⁵⁴ (SolarWinds, 2023)

Diese **SolarWinds-Tools** tragen dazu bei, dass Agenten und Benutzer schnelle Antworten finden und effizienter arbeiten können. Automatisierungseingebenen übernehmen penetrante einfache Alltagsaufgaben, wie Routing von Tickets, Sendung von Benachrichtigungen bei bestimmten Metadaten und Erstellung neuer Tickets bei Ablauf von Gerätewartungen. **Workflows und Runbooks** ermöglichen standardisierte Schritte für Serviceanfragen und Incident-Management festzulegen und sicherzustellen, dass Agenten diese Schritte genau befolgen können und verfügen auch über umfassende Anpassungsmöglichkeiten, mitsamt der Integration von Genehmigungsprozessen, Benachrichtigungen und externen Systemen. Durch die **Integration von KI in die ITSM-Lösungen** unterstützt SolarWinds Teams weltweit die Komplexität deren IT-Umgebungen zu reduzieren und sich auf die Bewältigung der wichtigsten Aufgaben zu konzentrieren.⁵⁵



Abbildung 52: künstliche Intelligenz für IT-Abläufe bei SolarWinds

Die Erweiterung des **SolarWinds® Service Desk** ermöglicht das Ticketvolumen zu reduzieren, indem User bei einfacher Problembewältigung unterstützt werden und so IT-Experten sich auf komplexere Aufgaben konzentrieren können.

Eine weitere Funktion ist der **AI-Virtual Agent**, um auf Benutzerfragen zu antworten und bei der Fehlerbehebung zu helfen durch die Nutzung von effizienten Servicebereitstellungen auf dem Service Desk mit über 200 Cloud-Anwendungen. Durch ständiges, kontinuierliches Lernen werden im Laufe der Zeit relevante Informationen und Lösungen angeboten.

Durch das **AI-gesteuerten ITSM-Upgrade** werden kontinuierliche Weiterentwicklungen von SolarWinds sichergestellt, dass bereits Cloud-basierte Beobachtungslösungen und ein neues Partnerprogramm integriert wurden. Die Kombination von Beobachtung und Service-Management in der SolarWinds-Plattform bietet eine einfache und sichere Lösungen für IT-Profis.⁵⁶

2.3.5.3. Demo von SolarWinds

SolarWinds Test-Demo unter [Book a demo](#) zu buchen.⁵⁷

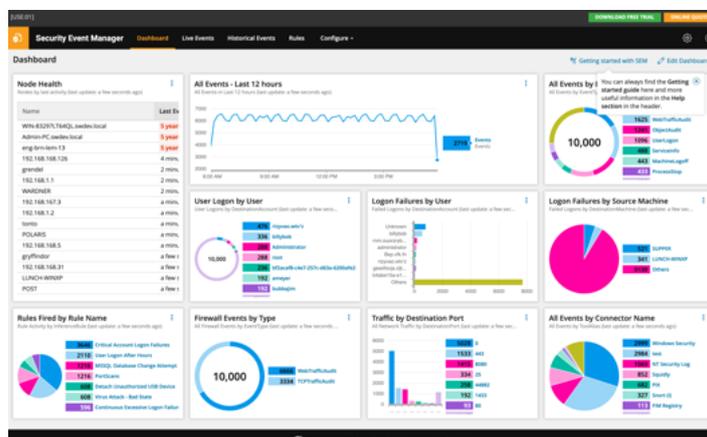


Abbildung 53: SolarWinds SIEM Demo

vgl.⁵⁵ (SolarWinds, 2024)

vgl.⁵⁶ (SolarWinds, 2023)

vgl.⁵⁷ (SolarWinds)

2.3.6. ManageEngine

Im folgenden Abschnitt 3.2.4.5. wird Tool ManageEngine etwas genauer beschrieben:^{58 59}

ManageEngine ist eine IT-Management-Division der privat geführten Zoho Corporation und konzentriert sich als Unternehmen mit seinen Produkten Applications Manager und Site24x7 auf On-Premises- und SaaS-Bereitstellungen für kleinen und mittelständischen Unternehmen (KMU). ManageEngine hat das Ziel, die kontinuierliche Weiterentwicklung seiner KI-Lösung Zia sowie eine vertiefte Integration in die Ereigniskorrelation und Sicherheitsaspekte zu verstärken.



Abbildung 54: ManageEngine

2.3.6.1. wichtige Funktionen von ManageEngine

Laut ManageEngine Informationsmaterial sind folgende Service angeboten:⁶⁰

Endpoint Central bedient automatische Endgeräte-Management mit Konfigurationen für das gesamten Netzwerkfunktionen, Schutz vor Vielzahl von Bedrohungen, Erstellung von Audits
Patch Management für über 1.000 Windows-, Mac-, Linux- und Drittanbieter-Anwendungen mit Automatisierung, Testen von vor der Bereitstellung, Integration von kritische Zero-Day-Patches.

Asset Management zur Überwachung von der gesamten Hardware und Software in Echtzeit, Compliance bei den Software-Lizenzen, analysieren von Software-Nutzungsstatistiken und Software-Metering sowie über 20 vordefinierten Berichten für Hardware, Software, Inventar und Lizenz-Compliance.

Software Development wird installiert/deinstalliert von MSI- und EXE-basierte Anwendungen, Bereitstellung von Software über das Self-Service-Portal, über 8.000 vordefinierte Vorlagen für die Bereitstellung von Anwendungen, sowie das Repository mit Software-Paketen und bestimmter Benutzeroptionsrechten.

Mobile Device Management automatisiert die Registrierung und Authentifizierung von mehreren BYOD- und Unternehmensgeräten, kontrolliert das Betriebssystem-Updates und der Problemebehebung bei mobilen Geräten, verwendet vorkonfigurierte und anpassbare Berichte über die mobilen Endgeräte des Unternehmens.

Mobile Application Management mit Konfigurationen nach Sicherheitsrichtlinien für WLAN, VPN, E-Mail etc. und Verhinderung von unbefugten Zugriffen auf geschäftliche E-Mails, sorgen für sichere Bereitstellung, Speicherung und Darstellung von Inhalten durch Verschlüsselung auf Geräteebene, sodass persönliche und unternehmenseigene Arbeitsbereiche auf BYOD-Geräten getrennt werden sowie aufspüren von verlegten Geräten, sperren diese und löschen darauf befindlichen Daten.

Mobile Security Management mit eigenes App-Repository, installieren, aktualisieren, konfigurieren, entfernen sowie verwalten von nur vertrauenswürdige Unternehmens-Apps mit Lizenzen und App-Berechtigungen, um bösartige/nichtssichere Apps auf eine schwarze Liste und verhindern, dass Benutzer Unternehmens-Apps deinstallieren.

vgl.⁵⁸ (ManageEngine)

vgl.⁵⁹ (ManageEngine, 2024)

vgl.⁶⁰ (ManageEngine)

OS-Bereitstellung mit intelligenten Online- und Offline-Imaging-Techniken, zentralen Repository und erfassten von Images mit Hilfe von Bereitstellungsvorlagen für verschiedene Rollen und Abteilungen im Unternehmen, problemlose Bereitstellung auf verschiedenen Hardware-Typen sowie Bereitstellung diverse Aktivitäten, wie die Installation von Anwendungen, die Konfiguration von Computereinstellungen und vieles mehr.

Endpoint Security Add-on (nur für On-Premises-Version) schützen Netzwerk vor Zero-Day-Schwachstellen durch kritische Bedrohungsanalysen und umfassendes Patch-Management, vor Browser vor Angriffen durch verwendeten Plugins, Add-ons und Erweiterungen überwachen, Überwachung und Einschränkung von USB-Geräten und anderen Peripheriegeräten, Datenübertragungen über BitLocker verschlüsselte Geräte, Blockierung sog. High-Risk-Anwendungen, durch eine Blacklist und Whitelist für Anwendungen, Verwaltung von End-point-Berechtigungen, Erkennung und Behebung von Ransomware-Infektionen mit Hilfe intelligenter Methoden zur Verhaltenserkennung (Behavior Detection) sowie Schützung von sensiblen Daten mit fortschrittlicher Data Loss Prevention vor unbefugter Datenweitergabe oder Datendiebstahl.

System-Tools überwacht und analysiert remote verwaltete Systeme, die durch das Anzeigen der Details zu den darauf laufenden Tasks und Prozessen, booten oder hochfahren der Rechner mit Wake-on-LAN aus der Ferne oder planen, Veröffentlichungen von Ankündigungen unternehmensweit oder nur für Techniker sowie planen von Defragmentierung, Überprüfung und Bereinigung von Festplatten für lokale oder entfernte Workstations

Fernsteuerung durch sichere Remote Control verschiedene Compliance-Vorschriften wie HIPAA, PCI DSS etc., Behebung von Problemen mit Remote-Desktops nahtlos und bei Bedarf durch die Zusammenarbeit mehrerer Benutzer - auch im Homeoffice, Integration von Video-, Anruf- und Chatfunktionen sowie die Optionen für die Dateiübertragung zwischen Rechnern, Fernsteuerungssitzungen zu Audit-Zwecken, Sperrung von Tastaturen und Mäuse von Endanwendern und verdunkeln deren Bildschirme, um die Vertraulichkeit während Remote-Control-Sitzungen zu gewährleisten sowie durch 128-Bit-AES-Verschlüsselungsprotokolle bei Fernsteuerungsvorgängen.

Konfigurationen von standardisierten von Desktop-, Computer-, Anwendungs- und Sicherheitseinstellungen mithilfe von Basiskonfigurationen; über 40 Konfigurationen für Benutzer und Computer oder Erstellung von eigenen Vorlagen für häufig verwendete Konfigurationen; über 300 Skripten im Script Repository; Einschränkung von Nutzung von USB-Geräten (z. B. Drucker, CD-Laufwerke, externe Geräte, Bluetooth-Geräte, Modems und andere Peripheriegeräte) im Netzwerk sowohl auf Benutzer- als auch auf Computerebene ein; effektives Energiemanagement; im Energieschemata; inaktive Computer abschalten; Berichte über die Systembetriebszeit anzeigen lassen; konfigurieren von Browser-, Firewall- und Sicherheitsrichtlinien und kontrollieren der Zugriffe auf Dateien, Ordner sowie die Registry mit Hilfe der Zugriffsrechteverwaltung sowie konfigurieren von Warnungen für in kurze ablaufende Passwörter und geringen Speicherplatz auf dem System.

Berichte verfügt über 200 sofort einsatzbereite Active-Directory-Berichte über Benutzer, Computer, Gruppen, Organisationseinheiten (OUs) und Domänen, Systembetriebszeit anzeigen; Benutzeranmeldung mit Benutzer- Anmeldeberichten oder zu Patches; Konfigurationen und Ereignisse für Audits.

2.3.6.2. ManageEngine mit KI

Neue KI von Zoho Corporation entwickelte KI-Engine wird in seinem gesamten Portfolio eingesetzt. Die Lösung ManageEngine APM unterstützt Teams dabei, Anomalien zu erkennen, prädiktive Analysen durchzuführen und Daten durch Berichte/Dashboards zu visualisieren, um detaillierte Einblicke in Leistungsprobleme zu gewinnen.

Preisgestaltung für KMUs und MSPs von ManageEngine zeichnet sich durch Wettbewerbsfähigkeit und eine einfache Implementierung aus, was sie besonders geeignet für kleinere Unternehmen mit begrenzten Budgets und Personalressourcen macht und ebenso sind diese für Organisationen geeignet, die im Modell des **Managed Service Providers** (MSP) Dienstleistungen für andere Unternehmen anbieten.

Breites Produktportfolio erweitert die Unternehmensfunktionalitäten über Application Performance Monitoring (APM) hinaus und umfasst IT-Betriebsfunktionen wie Netzwerküberwachung, Infrastrukturmanagement und den IT-Service-Desk.⁶¹

Malware-Prävention mit KI wird zur Malware-Erkennung eingesetzt, um unsere Hardware in der IT-Umgebung zu schützen. Diese analysieren Dateien, Bilder und andere herunterladbare Daten auf Malware. Unser System überwacht Anhänge auf unserem Server und prüft eingehende Dateien auf verdächtige Skripte, insbesondere bei PDF- und exe-Dateien. Es extrahiert und analysiert Merkmale, um bösartige Daten zu identifizieren. Diese Analyse erfolgt in zwei Phasen: statische Analyse, die Dateien auf verdächtige Aktivitäten und Metadaten überprüft, und dynamische Analyse, die den Code auf bösartiges Verhalten überwacht.

Statische Analyse sammelt im IT- System zunächst Informationen über Malware, ohne den Code anzusehen und dabei unterscheidet es zwischen dem erwarteten Verhalten einer normalen Datei und dem potenziell schädlichen Verhalten einer Malware, wie z.B. wiederholte Aktivitäten oder unlesbare URLs. Die Extraktion und Analyse von Metadaten wie Dateiname, Typ und Größe liefern Hinweise auf die Art der Malware. Weiterhin analysiert das IT-System den Code, um böswillige Absichten zu erkennen, ohne ihn ausführen zu müssen. Dennoch kann manche Malware durch statische Analyse unentdeckt bleiben, insbesondere wenn sie sehr raffiniert ist. In solchen Fällen ist eine dynamische Analyse erforderlich.

Dynamische Analysen führen im IT-System die Dateien in einer geschützten Blackbox-Umgebung aus, um Malware zu identifizieren. Diese Umgebung isoliert potenzielle Schäden von der Produktionsumgebung und kann nach der Analyse zurückgesetzt werden. Durch Überwachung der Malware wird die Umgebung modifiziert, z.B. durch Änderungen an Registry-Schlüsseln oder Kommunikation mit externen Hackern.

Analyse des Benutzer- und Entitätsverhaltens in dieser Phase führt unser System die Dateien in einer geschützten Blackbox-Umgebung aus und identifiziert dabei Malware. Diese virtuelle Umgebung ist isoliert vom Rest des Netzwerks und verhindert potenzielle Schäden in der Produktionsumgebung. Nach der Analyse kann die Blackbox ohne dauerhafte Schäden zurückgesetzt werden. Das KI-System überwacht die Blackbox, um Änderungen durch die Malware zu erkennen, wie z.B. an Registry-Schlüsseln, IP-Adressen, Domainnamen oder Dateipfaden, sowie Kommunikation mit externen Hackern.

ManageEngine umfasst verschiedene Arten von Entitäten mit einzigartiges Verhalten, das als "normal" betrachtet wird. Abweichungen davon können Sicherheitsbedrohungen darstellen. Das KI-System überwacht Benutzer- und Entitätsverhalten, um Kontoübernahmen, interne Bedrohungen und Datenabfluss zu erkennen und wird als User-Entity- und Verhaltensanalyse (UEBA) bezeichnet. Das System wertet zukünftiges Verhalten basierend auf einem normalen Zustand aus und löst bei Abweichungen Anomalien aus und führt präventive Maßnahmen basierend auf diesem Verhalten durch.



vgl.⁶¹ (ManageEngine)

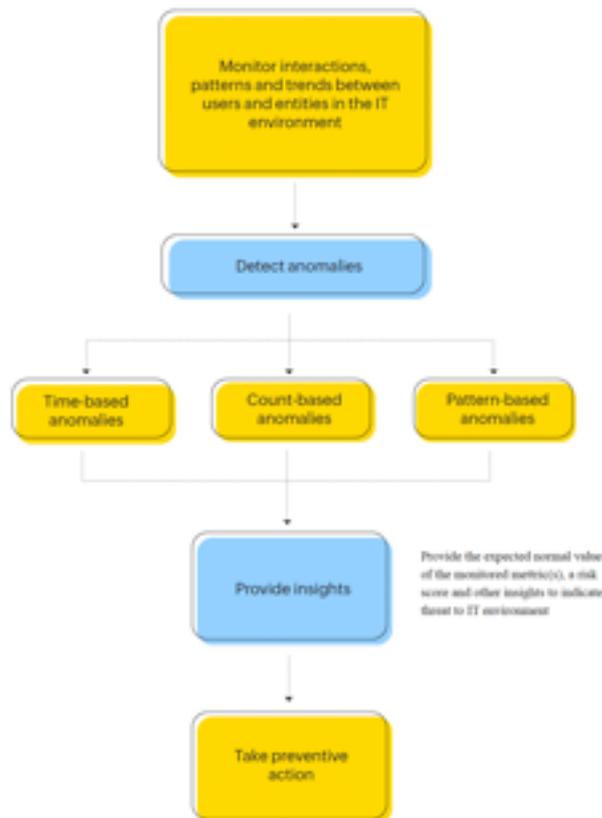


Abbildung 55: Analyse des Benutzer- und Entitätsverhaltens mithilfe von KI und Prozessflussdiagramm für die Analyse von Benutzerentitäten und -verhalten

Verhaltensmuster des KI-System überwacht Benutzer-Entitäts-Interaktionen, Muster und Trends, um Abweichungen vom Normalen zu erkennen basierend auf diesem normalen Verhalten erkannt und in drei Kategorien eingeteilt:

Häufige Muster werden nach dem Einloggen überprüft der Benutzer sofort E-Mails und Nachrichten bevor er mit anderen Aufgaben fortfährt.

Seltene Muster unmittelbar nach dem Einloggen überprüft der Benutzer E-Mails und Nachrichten und geht dann zu anderen Aufgaben über, jedoch nicht täglich, sondern gelegentlich.

Ungesehene Muster von Benutzer greifen nach dem Einloggen direkt auf einen sicheren Bereich der IT-Umgebung zu, anstatt E-Mails zu überprüfen und sind Warnzeichen und haben hohe Priorität, daher sendet unser System sofort einen Alarm.

Zeitbasierte Anomalien sind Abweichungen bei Benutzeranmeldungen. Nachrichten werden an IT-Administratoren weitergeleitet.

Zählbasierte Anomalien sind Abweichungen bei Dateizugriffen eines Benutzers.

Musterbasierte Anomalien sind Abweichungen von üblichen Anmeldezeiten und können auf Insider-Bedrohungen oder Datenexfiltration hinweisen. IT-Admins können das System anpassen.

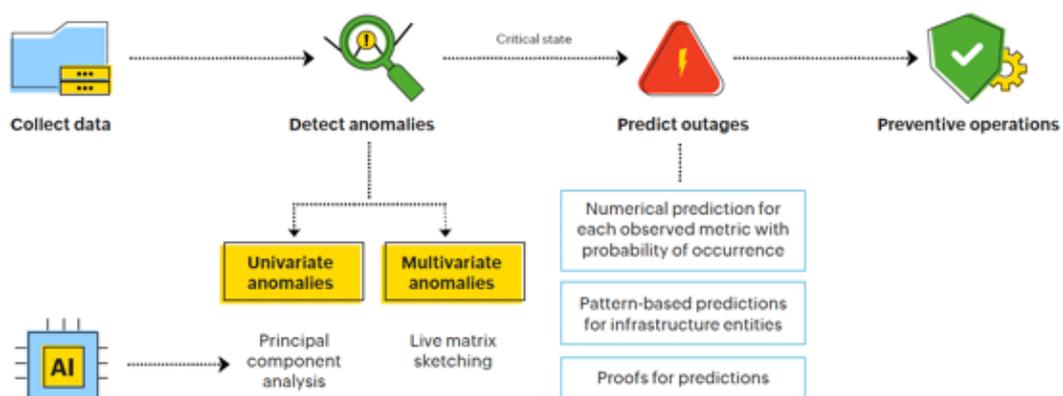


Abbildung 56: Prozessflussdiagramm für die Vorhersage von Ausfällen

Das KI-System passt sich automatisch an verändernde Datenmuster an und markiert Anomalien. IT-Admins können sich auf präventive Maßnahmen konzentrieren und liefert Erklärungen zu Anomalien, um fundierte Entscheidungen zu ermöglichen.⁶²

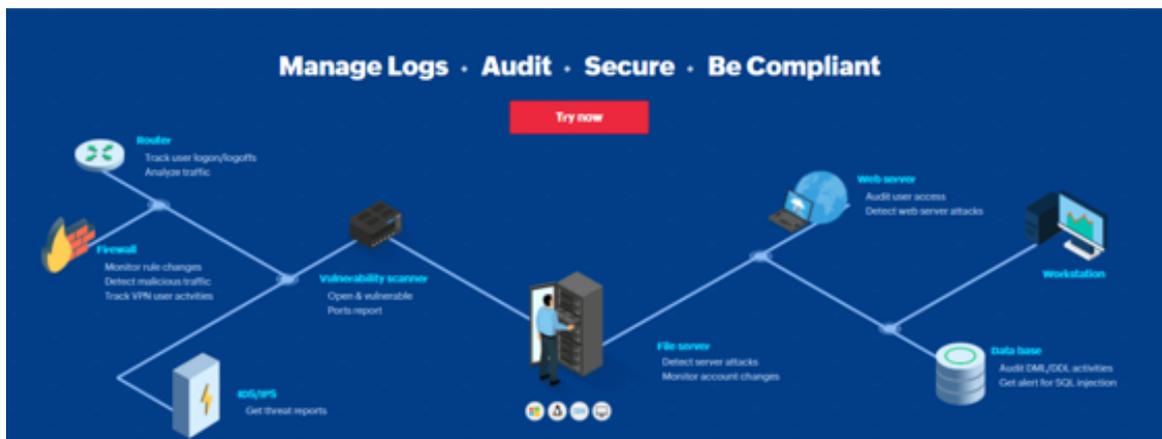


Abbildung 57: MangleLogs - Audit - Secure – Be Compliant

2.3.6.3. Demo von ManageEngine

ManageEngine Test-Demo unter [Book a demo](#) zu buchen.⁶³

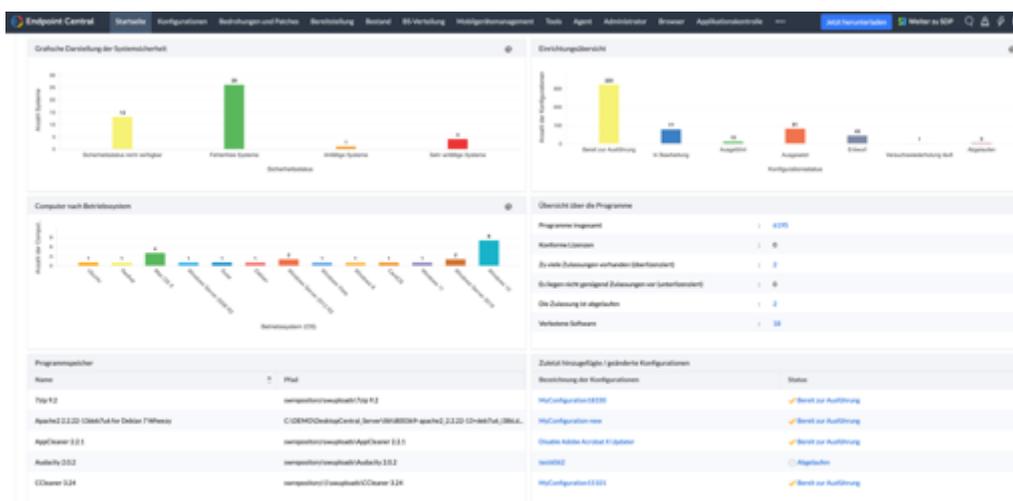


Abbildung 58: ManageEngine Demo

2.3.7. Splunk

Im folgenden Abschnitt 3.2.4.6. wird das Tool Splunk etwas genauer beschrieben.^{64 65}

Splunk als Observability Cloud bietet umfassende Abdeckung in verschiedenen Observability-Bereichen, darunter Infrastruktur, **Application Performance Monitoring (APM)**, **Digital Experience Monitoring (DEM)**, **AIOps** und **Incident Intelligence**. Die betrieblichen Aktivitäten von Splunk sind diversifiziert, und das Kundensegment besteht überwiegend aus großen Unternehmen. Splunks beinhaltet die Implementierung von KI-gesteuerter End-to-End-Full-Stack-Observability für hybride Umgebungen, die Integration von Observability und Sicherheitslösungen sowie die Bereitstellung von Sichtbarkeit und Kontrolle über die Nutzung und Kosten der Metrikenplattform.

vgl.⁶² (ManageEngine)

⁶³ (ManagerEngine)

vgl.⁶⁴ (Splunk)

vgl.⁶⁵ (Splunk)

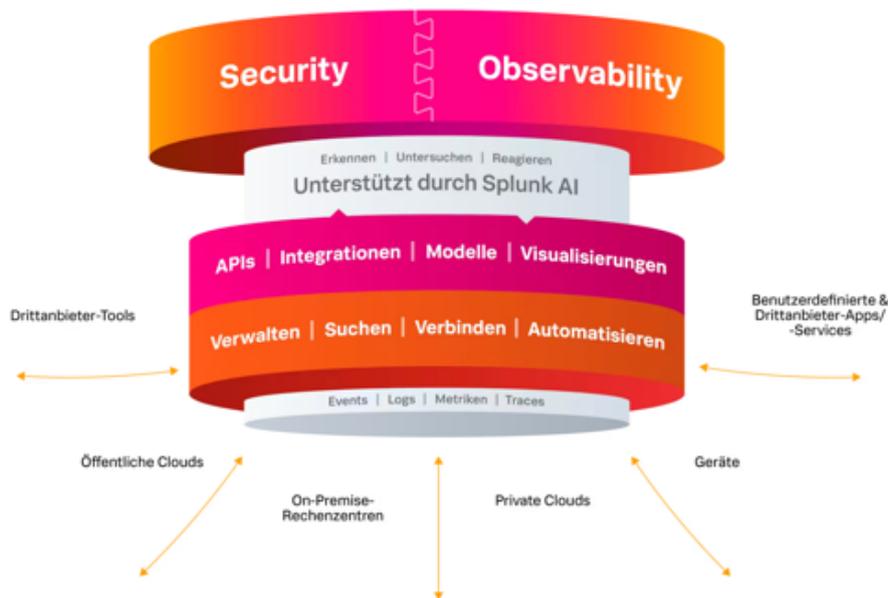


Abbildung 59: Splunk Observability Cloud Schema

2.3.7.1. wichtige Funktionen von Splunk mit KI

OpenTelemetry-Unterstützung bei Splunk Observability Cloud ist OpenTelemetry-nativ und ermöglicht die Erfassung und Analyse von OTEL ohne den Bedarf an proprietären Agenten und macht Splunk APM besonders effektiv für cloud-native, servicebasierte und **Mesh-App- und Service-Architekturen (MASA)**.

Erweiterung der Abdeckung der dominanten Anbieter für Log-Monitoring und SIEM in großen Unternehmen bietet Splunk seinen bestehenden Kunden die Möglichkeit, Observability-Lücken zu schließen sowie trägt dazu bei der Vereinfachung von Beschaffungsprozessen und bei der Reduzierung der Anzahl externer Vertragspartner.

Skalierbarkeit basiert auf einer hochelastischen Mandantenarchitektur, die eine automatische Skalierbarkeit ermöglicht und dabei können Kunden bis zu 3x über die Abonnementrate hinaus skalieren, um elastische Anforderungen zu unterstützen sowie eine anpassbare Ratenbeschränkung pro Kunde gewährleisten und somit eine effiziente Ressourcennutzung.

End-to-End-Überwachung bietet eine umfassende Überwachung von Anwendungen, Infrastruktur und Benutzererfahrung und ermöglicht Performance-Probleme schnell zu identifizieren und zu beheben.

Distributed Tracing unterstützt verteiltes Tracing, was bedeutet, dass der Weg von Anfragen durch verschiedene Komponenten der Anwendungen verfolgen können und somit besonders nützlich in Mikrodienstarchitektur ist.

Metriken und Protokollierung ermöglicht das Sammeln und Analysieren von Metriken und Protokollen aus verschiedenen Quellen und hilft bei der Identifizierung von Leistungsproblemen und Fehlern.

Real-Time-Monitoring bietet Echtzeitüberwachungsfunktionen, die es ermöglichen, sofort auf auftretende Probleme zu reagieren und Engpässe in der Leistung zu identifizieren.

Korrelation von Daten ermöglicht die Korrelation von Daten aus verschiedenen Quellen, was es erleichtert, Zusammenhänge zwischen verschiedenen Komponenten und Systemen zu erkennen.

Automatisierte Warnungen ermöglichen das Einrichten von Warnungen für bestimmte Ereignisse und Leistungsparameter und Benutzer erhalten Benachrichtigungen, wenn festgelegte Schwellenwerte überschritten werden.

Integration mit anderen Tools kann nahtlos mit anderen Tools und Plattformen integriert werden, um Daten aus verschiedenen Quellen zu aggregieren und zu analysieren.

Analytics und Dashboards bieten leistungsstarke Analysefunktionen und die Möglichkeit, benutzerdefinierte Dashboards zu erstellen.

Key Performance Indicators (KPIs) überwachen und gewinnen Einblicke in die Systemleistung.

Security Monitoring bietet die Funktionen zur Sicherheitsüberwachung, um potenzielle Sicherheitsbedrohungen zu erkennen und darauf zu reagieren.⁶⁶



Abbildung 60: Splunk Observability

Schnelleres Troubleshooting mit Log Observer Connect erweitert die Anwendungsfälle Ihrer Splunk-Cloud-Protokolle für schnelles Anwendungs- und Infrastruktur-Debugging in einer intuitiven Low-Code-Oberfläche.

Dashboards und Visualisierungen sind anpassbare Dashboards und Visualisierungen ermöglichen klare Datenpräsentation für verschiedene Zielgruppen.

Monitoring und Benachrichtigungen sind durchgängige Monitoring von Events, Umgebungsbedingungen und KPIs mit regelmäßigen Suchläufen und benutzerdefinierten Alert-Aktionen für Echtzeit-Warmmeldungen.

Reporting wird mit Echtzeit-Berichte für verschiedene Intervalle erstellt und in sicheren Formaten wie PDF abgespeichert.



Abbildung 61: Quellen von Splunk

Traces sind präzise Abschnitte der Benutzerreise, die ausführliche Informationen über welche Dienste aufgerufen wurden, auf welchen Containern, Hosts oder Instanzen sie ausgeführt wurden und welche Ergebnisse jeder einzelne Aufruf erzielt hat. Durch granulare Traces lassen sich die Interaktionen und Abläufe innerhalb komplexer Systeme verfolgen.

vgl.⁶⁶ (Splunk)

Frameworks wie OpenTelemetry ermöglichen die Erfassung von Metriken und Traces aus Anwendungen in verschiedenen Programmiersprachen. Weitere Tools wie **collectd** für Metrik-Sammlungen, **statsd** für Listenstatistiken, **fluentd** für Log-Datensammlungen, **Zipkin** und **Jaeger** welche als Open-Source-Back-End-Systeme für verteiltes Tracing fungieren.

Logs (Ereignisse) können in verschiedenen Formaten auftreten in Form als einfacher Text, strukturiert oder binär. Die beeindruckende Bandbreite an Ereignisquellen reicht von System- und Server-Logs (wie **syslog** und **journald**) über Firewalls und Intrusion Detection Systems bis hin zu sozialen Medienfeeds (wie Twitter sowie umfasst Anwendungs-, Plattform- und Server-Logs (wie log4j, log4net, Apache, MySQL, AWS).

Metriken für Observabilität, definiert von renommierten Analysten wie Gartner, Forrester und IDC, sind entscheidend für eine umfassende Observabilität. Dazu gehören **Events** (Logs), **Traces** und **Metriken** aus verschiedenen Quellen wie System- und Server-Logs, Web-Tracking-Skripte und Geschäftsmetriken. Die effektive Integration und Analyse dieser Daten ermöglicht ein detailliertes Verständnis der Systeme und ihrer Leistung. Die Nutzung von Metrikdaten für verbesserte Such-Performance und kosteneffiziente Speicherung. Die Quellen für Metriken sind vielfältig und umfassen Systemmetriken wie CPU-Auslastung, Speicherverbrauch und Festplattennutzung.

Analytics Workspace bietet visuelle Analysefunktionen ohne SPL-Kenntnisse, wie vielfältig und umfassen Systemmetriken wie CPU-Auslastung, Speicherverbrauch und Festplattennutzung, wie Infrastrukturmetriken durch AWS CloudWatch; Web-Tracking-Skripte wie Google Analytics und Digital Experience Management die Verfolgung von Benutzerinteraktionen, während Anwendungsagenten und -sammler (wie APM und Fehlerverfolgungstools); Geschäftsmetriken, darunter Umsatz, Kundenanmeldungen, Absprungraten und der Warenkorbabbruch, liefern wertvolle Einblicke in den geschäftlichen Erfolg.

Moderne Techniken der Ereignisverarbeitung spielen eine Schlüsselrolle bei der Umwandlung von Daten in Erkenntnisse für umfassende Observabilität. Diese Techniken bieten geteilte Erkenntnisse, kollaborative Reaktionen auf Vorfälle, datengestützte Entwicklung und intelligente Betriebsabläufe. Ein effektives System zur Ereignisverarbeitung sollte Kontext hinzufügen, Analysen durchführen, Daten entdoppeln und alle Daten sammeln können.

Die **umfassende Sammlung von Observabilitäts- und Überwachungsdaten** ist entscheidend, da verschiedene Datenquellen gesammelt werden, darunter Netzwerkflussdaten, virtuelle Server-Logs, Cloud-Dienste, Docker-Informationen und Container- und Mikroservice-Architekturen, Drittanbieterdienste, Kontrollsysteme, Dev-Automatisierung, Infrastruktur-Orchestrierung, Signale von mobilen Geräten, Kennzahlen für Business Analytics, Signale aus der sozialen Sentiment-Analyse, Kundenbindung und Erfahrungsanalyse, Nachrichtenbusse und Middleware, bieten Einblicke in die Leistung von Systemen. Die Integration und Analyse dieser Datenquellen ermöglicht Unternehmen ein umfassendes Bild ihrer IT-Infrastruktur und Anwendungslandschaft, was die Grundlage für erfolgreiche Observabilität schafft.

Machine Learning-Toolkit (ML-TK) sind integrierte Machine-Learning-Analysen und die Möglichkeit, eigene ML-Modelle einzurichten, bieten flexible Lösungen für unterschiedliche Use Cases.

Skalieren und Verwalten von Datenvorhaltung kann passgenau konfiguriert und Speicherkapazität kann einfach nach Bedarf erworben werden.

Dynamische Datenablage durch Dynamic-Data-Optionen ermöglichen die Anpassung an Langzeitarchivierungsanforderungen unter Berücksichtigung von Auditing- und Compliance-Vorgaben.

Integrationen mit Splunk-Berichte können in verschiedene Anwendungen eingebunden oder per ODBC-Integration in Microsoft Excel, Tableau usw. verwendet werden.

Benutzerauthentifizierung von Splunk Cloud Plattform unterstützt Single Sign-on per SAML über gängige Identitätsanbieter und kann mit verschiedenen Authentifizierungssystemen integriert werden.⁶⁷

Durch die **Verwendung KI und ML** ermöglichen integrierte lernfähige Algorithmen genaue prädiktive Analyse und so zukünftige Ereignisse vorherzusagen und durch das maschinelle Lernmodelle eine genaue Perspektiven auf historische und Echtzeitdaten zu entwickeln und

vgl.⁶⁷ (Splunk)

so gesteuerte Analytik reduziert von Ereignischaos sowie falschen Positiven durch multivariate Anomalieerkennung durch KI-Algorithmen eine präzisere Identifizierung ungewöhnlicher Muster und Abweichungen, wodurch unnötige Alarme minimiert werden; **automatisches Verbergen von Duplikatereignissen** mit automatische Unterdrückung von Duplikatereignissen können relevante Vorfälle hervorgehoben und Alarmstürme reduziert werden; **leichtes Durchsuchen von großen Mengen von Ereignissen** mithilfe von Filterung, Markierung und Sortierung können Anwender mühelos durch umfangreiche Datensätze navigieren und sich auf wesentliche Informationen konzentrieren sowie **anreichern und hinzufügen von Kontext zu Ereignissen** durch KI-gesteuerte Analytik ermöglicht die automatische Anreicherung von Ereignissen mit relevantem Kontext, wodurch sie informativer und besser handhabbar werden.^{68 69} Die Einbindung von Künstlicher Intelligenz (KI) und Maschinellem Lernen (ML) in Observabilitätssysteme ist entscheidend bei der Maximierung des Werts der gesammelten Daten. Durch diese Integration wird es möglich, komplexe Zusammenhänge besser zu verstehen und fundierte, prädiktive Einblicke in die zukünftige Leistung und Sicherheit von geschäfts-kritischen Systemen zu gewinnen.

2.3.7.2. Splunk mit KI

Splunk bietet zwei Möglichkeiten, KI und ML zu nutzen: **Out-of-the-Box-Funktionen** für direkt integrierte Workflows und kundenspezifische Anpassungen. Die Funktionen umfassen Anomalieerkennung, Prognosen, Vorhersagen und Datenclustering. Diese können entweder über Assistenten oder direkt mit der Splunk-Suchsprache SPL genutzt werden. Splunk bietet auch ML-basierte Tools in verschiedenen Produkten wie Splunk Enterprise Security, Splunk User Behavior Analytics und Splunk IT Service Intelligence. Zusätzlich gibt es unterstützende KI-Tools wie den Splunk AI Assistant und die Splunk App for Anomaly Detection. Spezielle Add-ons wie das **Machine Learning Toolkit (MLTK)** und die Splunk App for **Data Science and Deep Learning (DSDL)** ermöglichen die Ausführung von ML-Workloads und die Integration fortschrittlicher ML- und Deep-Learning-Systeme in die Splunk-Plattform.

Observability ist ein moderner Monitoring-Ansatz, der umfassende Transparenz und Kontextinformationen über den gesamten Infrastruktur-Stack, alle Anwendungen und die Customer Experience bietet. Splunk wurde von TrustRadius für Ereignisanalysen ausgezeichnet und von GigaOm als Leader für Cloud-Observability sowie von Gartner als Leader in den Bereichen **Application Performance Monitoring (APM)** und Observability anerkannt. Diese Erfolge basieren auf Splunks zentraler Datenplattform, die es ermöglicht, Logs, Kennzahlen und Trace-Daten nahezu in Echtzeit abzurufen und zu visualisieren.

Künstliche Intelligenz für IT Operations (AIOps) ist ein fortgeschrittenes Konzept für Observability, Incident Management und Incident Response.

Die **Splunk AIOps-Lösung IT Service Intelligence** verwendet ML für adaptive Schwellenwerte, Anomalieerkennung, prädiktive Analysen und den Smart Mode zum automatischen Gruppieren von Warnungen. Eine weitere wichtige Anwendungsmöglichkeit von ML ist die Überwachung komplexer Systeme über die menschliche Ebene hinaus, ermöglicht durch Splunk Infrastructure Monitoring. Funktionen wie Assistenten zur Anomalieerkennung und zur Vorhersage der Einrichtung von Warnungen werden durch diese ML-Funktionen unterstützt. Splunk empfiehlt eine kurze Auswirkungsbewertung vor jedem neuen KI-Projekt, um eine erfolgreiche Implementierung sicherzustellen. Dies würde zumindest abdecken, was das Projekt erreichen will, welche Risiken bestehen und wie gut es sich umsetzen lässt. Erfolgreiche ML-Projekte haben detaillierte, genau definierte Ziele – beispielsweise die Verbesserung der Erkennungsgenauigkeit oder die Verringerung des manuellen Verarbeitungsaufwands. KI und ML bergen unter anderem Risiken für Transparenz, Kontrolle und Fehlertoleranz. Unternehmen sollten daher kritisch darüber nachdenken, wie sie diese Risiken angehen. Die richtige Wahl der Splunk-Produkte kann Datentransparenz und -kontrolle gewährleisten. Die folgende Matrix erklärt, wie verschiedene Überlegungen zu Datentransparenz und -kontrolle die Auswahl von Splunk-Produkten beeinflussen können.

vgl.⁶⁸ (Splunk)

vgl.⁶⁹ (Splunk)

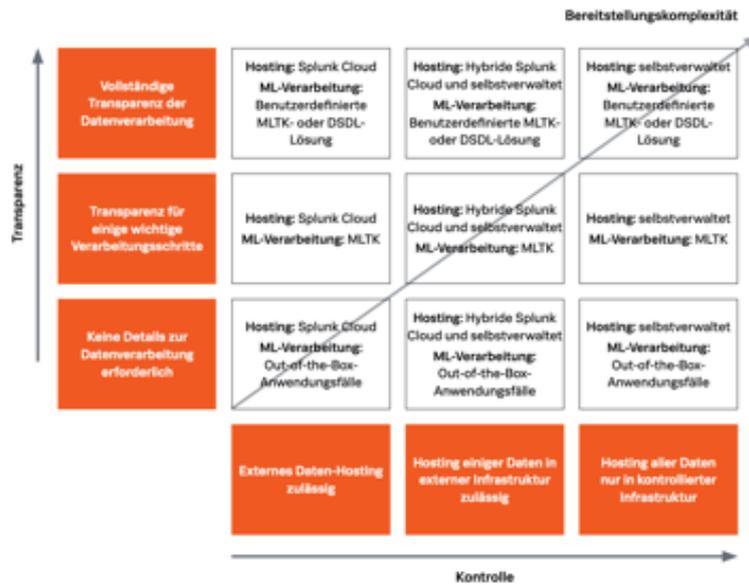


Abbildung 62: Risikobewertung bei MLTK- oder Out-of-the-Box-Anwendungsfällen

ML für Observability wird in verschiedenen IT- und OT-Bereichen eingesetzt, sodass jeder Anwendungsfall geschäftliche Herausforderung umfassen, den Lösungsansatz von Splunk und die Vorteile der Implementierung bereitstellen. Am meisten verwendeten ML-basierten Analysen in Splunk sind drei gängige Techniken: Anomalieerkennung, prädiktive Analysen und Clustering.



Abbildung 63: ML-basierten Analysen in Splunk

Prognose der Ressourcenauslastung mit ML-TK können Anwender Ressourcenauslastungsprognosen erstellen, z.B. durch geführte Workflows oder direkt in der Suchleiste. Durch das Infrastructure-Monitoring werden Warnmeldungen für kritische Auslastungen eingerichtet, welche basierend auf dem Prognoseverfahren der doppelten exponentiellen Glättung. Dabei erfordern Prognosen keine speziellen Data-Science-Kenntnisse mehr. Vorteile sind Vermeidung von potenziellen Ausfällen sind weniger manueller Aufwand für IT-Teams, eine bessere Vorbereitung und ein proaktives Handeln, um arbeitsintensive Ausfälle zu vermeiden und die Effizienz zu steigern.

Erkennung von Service-Performance-Problemen durch Common Information Model erstellt eine konsistente Normalisierung von Kennzahlen für verschiedene Infrastrukturen und Anwendungen. Die Splunk-ITSI ermöglicht dabei das Festlegen adaptiver Schwellenwerte für KPIs wie Anwendungsfehlerraten und die Vorhersage späterer Servicezustände und bietet ebenfalls ML-Mechanismen zur Identifizierung abweichender Kennwerte, zur verbesserten Überwachung der Anwendungsleistung, zur Vermeidung potenzieller Ausfälle und zur Beeinträchtigungen der Customer Experience sowie ermöglicht weniger manueller Aufwand durch die automatisierte Erkennung von Anomalien und entlastet dabei die IT-Experten und spart Zeit ein.

Monitoring der User Experience mit dem ML-TK können Benutzer zu einem besseren Verständnis der UX erlangen, um z.B. Ausreißer in Transaktionszeiten zu identifizieren oder Clients-basierte Attribute in Clustern zusammenzufassen, um abweichende Muster heraus-

zufiltern, um so eine höhere Erkennungseffizienz durch ML-gestützte UX-Referenzwerte in den Bereichen mit hohem Verbesserungspotenzial zu analysieren sowie eine frühzeitige Erkennung von Kapazitätsengpässen und UX-Problemen für die die User Experience zu verbessern.

Reduktion von Warnmeldungen reduzieren die Ansammlungen von Benachrichtigungen durch intelligentere Analysen. Dabei hilft Splunk-ITSI mit der Smart Mode-Funktion, Warnmeldungen automatisch zu gruppieren und manuelle Überprüfungen zu reduzieren. Es werden weniger manuelle Aufwände für Analysten betrieben, sodass eine effiziente und verbesserte Arbeitsweise erfolgt bzw. zu einer Reduzierung menschlicher Fehlerquote und einer minimierten Übersehensrate von wichtigen Vorfällen.

Vorhersage von Datenausfällen erfasst Daten aus verschiedenen Systemen und erstellt ML-Modelle, um die erwartete Anzahl von Ereignissen für jede Host-Datenquelle-Kombination zu hervorzuheben. Dabei überwacht ein Modell kontinuierlich die Datenfeeds und erkennt automatisch Abweichungen vom Sollwert, was zum Erkennen und Beheben von Anomalien während der Datenaufnahme zu einer hohen Datenverfügbarkeit beiträgt und dabei die betriebliche Effizienz erhöht.

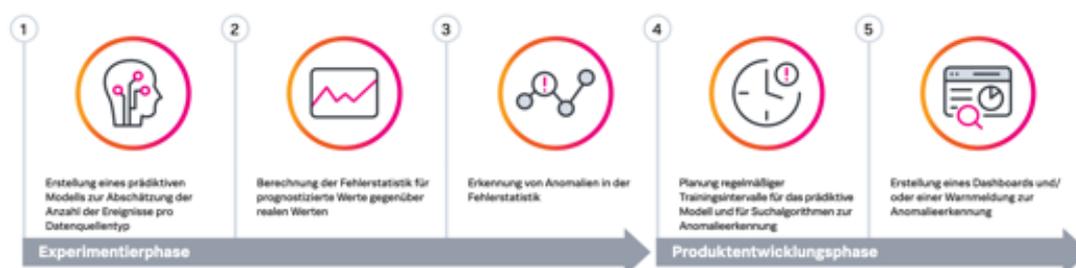


Abbildung 64: Vorhersage von Datenausfällen in Splunk

Monitoring von Mobilfunkmasten bietet mit dem ML-TK-Training und den Einsatz von ML-Modellen die Erkennung von Netzwerk-Traffic-Anomalien, zur Vorhersage von Traffic-Mustern und zur Gruppierung von Entitäten, die über geführte Workflows oder direkte Suchfunktionen von Telekommunikationsanbieter vielfältig verwendet werden können, um lokale Netzüberlastungen zu vermeiden, Datengeschwindigkeiten zu optimieren, Kundenbeschwerden zu reduzieren und die User Experience zu verbessern sowie die automatische Erkennung von Netzproblemen eine effizientere Nutzung von Netzressourcen das technische IT-Teams zu entlasten.⁷⁰

2.3.7.3. Demo von Splunk

Splunk Test-Demo unter [Book a demo](#) zu buchen.⁷¹

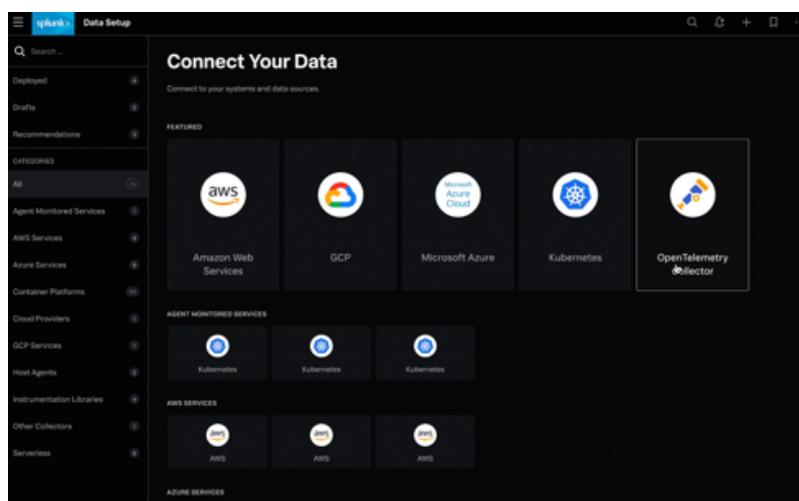


Abbildung 65: Splunk Demo

vgl.⁷⁰ (Splunk)

⁷¹ (Splunk)

2.3.8. Vergleiche der vorgestellten Tools

Die **Benutzeroberflächen der verschiedenen SIEM-Tools** bieten jeweils eigene Vorteile: **Splunk** überzeugt mit einer besonders benutzerfreundlichen Oberfläche, die mit individuell anpassbarer Dashboards eine hohe Übersichtlichkeit gewährleistet. Diese Anpassungsmöglichkeiten sind nahezu unbegrenzt, sodass Nutzer die Oberfläche nach ihren Bedürfnissen selbst gestalten können. **ManageEngine** bietet eine klar strukturierte, einfache Benutzeroberfläche, während **SolarWinds** und **Logpoint** durch eine gut übersichtliche Gestaltung punkten, die ebenfalls **Anpassungsmöglichkeiten** erlaubt. **LogRhythm** besticht durch eine **intuitive Bedienung** mit vielen Optionen zur individuellen Anpassung, während **IBM QRadar** eine robuste, umfassende **Benutzeroberfläche** mit zahlreichen Anpassungsmöglichkeiten bietet.

Die **Integration von Sysmon-Daten** (Systemmonitor für Windows-Systemdienste) wird von allen Tools gut unterstützt. **Splunk, ManageEngine, SolarWinds, LogRhythm, Logpoint und IBM QRadar** erlauben die **Einbindung von Sysmon-Logs** und bieten damit eine umfassende Analyse der Logereignisse. Alle Tools stellen die Logereignisse detailliert dar, wobei die Darstellungsweise je nach Tool unterschiedlich ausfällt. Besonders **starke Filterfunktionen** zur präzisen Analyse der Logs finden sich bei **Splunk, LogRhythm und IBM QRadar**.

Die **Erstellung fortgeschrittener Sicherheitsregeln** sowie umfassende **Kalibrierungsmöglichkeiten** sind bei **Splunk, LogRhythm und IBM QRadar** verfügbar. Diese Tools bieten effektive Mechanismen, um Sicherheitsregeln umzusetzen und Bedrohungen frühzeitig zu erkennen. Klare Alarme und eine übersichtliche Darstellung sorgen dafür, dass Teams über sicherheitsrelevante Vorfälle informiert werden, wobei **Echtzeit-Benachrichtigungen** bei **Splunk, LogRhythm und IBM QRadar** besonders hervorgehoben werden. Diese Tools gewährleisten zudem eine sofortige **Alarmierung bei sicherheitskritischen Ereignissen**.

In Bezug auf **Agenten und Forwarder** zeichnen sich alle Tools durch ihre Zuverlässigkeit und Robustheit aus. **Splunk, LogRhythm und IBM QRadar** bewältigen auch **große Datenmengen** effizient. Der Einstieg in die **Bedienung** fällt bei **ManageEngine und SolarWinds** etwas leichter, während **Splunk und IBM QRadar** eine **etwas längere Einarbeitungszeit** erfordern. Umfangreiche **Support- und Dokumentationsressourcen** stehen bei **Splunk, LogRhythm und IBM QRadar** zur Verfügung, um Nutzern bei der Einarbeitung und bei Problemen zu helfen.

Die **Implementierung der Tools** erfordert zwar überall Zeit, aber **Splunk** könnte aufgrund seines breiten Funktionsumfangs **mehr Ressourcen** für eine erfolgreiche Einführung beanspruchen. **Zusätzliche Features** wie leistungsstarke Dashboards, die Integration von KI zur Erkennung fortschrittlicher Bedrohungen sowie umfangreiche Bibliotheken mit Apps und Add-ons sind bei **Splunk, LogRhythm und IBM QRadar** zu finden. **IBM QRadar** wird oft als **Next-Generation-SIEM** bezeichnet, da es fortschrittliche Sicherheitsanalysemöglichkeiten bietet. Auch die anderen Tools entwickeln sich weiter, um den ständig wachsenden Bedrohungsszenarien gerecht zu werden.

Die Entscheidung für ein bestimmtes SIEM-Tool hängt letztlich von den individuellen Anforderungen, dem Budget und den Vorlieben der Organisation ab. **Splunk, LogRhythm und IBM QRadar** zählen zu den **führenden Lösungen**, unterscheiden sich jedoch in ihren **Schwerpunkten und Preisstrukturen**. **ManageEngine und SolarWinds** bieten gute Alternativen für Organisationen mit und **spezifischen Anforderungen begrenztem Budget**. **Logpoint** ist eine solide Wahl, die **Benutzerfreundlichkeit und starke Funktionen** vereint. Am Ende sollte die Entscheidung darauf basieren, welches Tool die Bedürfnisse der jeweiligen Organisation am besten erfüllt.

3. KI Anwendungen Monitoring

3.1. KI Grundlagen

Künstliche Intelligenz (KI) (*eng.* Artificial Intelligence (AI)) sind Maschinen oder Computer, die selbstständig agieren können mithilfe verschiedener Technologien, Algorithmen und Systeme, die die menschliche Intelligenz nutzen, um zu lernen, Aufgaben oder Probleme zu lösen, Muster zu erkennen und Entscheidungen zu treffen.

3.1.1. KI Haupttypen

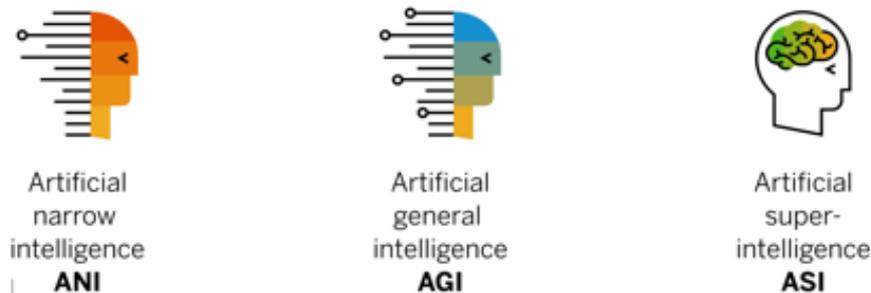


Abbildung 66: drei wichtigsten Arten von KI

Schwache künstliche Intelligenz (*engl.* Artificial narrow intelligence (ANI) bezeichnet eine schwache KI, die trotz komplexer Algorithmen und neuronaler Netzwerke auf spezifische Aufgaben beschränkt ist, wie z.B. Gesichtserkennung und selbstfahrende Autos. Der Begriff "**schwach**" bezieht sich nicht auf ihre Leistungsfähigkeit, sondern darauf, dass sie noch nicht menschliche Kompetenzen erreicht haben.

Starke künstliche Intelligenz (*engl.* Artificial general intelligence (AGI)), bezeichnet eine **starke KI**, soll jede intellektuelle Aufgabe wie ein Mensch bewältigen können. Im Unterschied zu schwachen KI können starke KI-Systeme auf nicht vorab definierte Aufgaben und Situationen reagieren, wie z.B. der Summit Superhochleistungsrechner hat die Leistungskraft, extrem schnelle Berechnungen durchzuführen, wie z.B. 200 Billionen Berechnungen in der Sekunde anstatt für 1 Milliarde Jahre Menschendenken.

Künstliche Superintelligenz (*engl.* Artificial general intelligence (ASI)), bezeichnet eine **sehr starke KI**, besitzt theoretisch ein Bewusstsein für ihr Dasein und übertrifft menschliche Fähigkeiten bei Verarbeitung und Analyse deutlich. Obwohl eine solche Welt unwahrscheinlich erscheint, müssen ethische Richtlinien zu beachtet werden, da die KI sich rapide weiterentwickelt und uns in vielen Aspekten übertreffen könnte, wie von Stephen Hawking betont.⁷²

Hinweis: Je nach Literatur können auch weitere KI-Haupttypen betrachtet werden.

3.1.2. KI Technologien

Die Technologien, die KI ermöglichen, umfassen maschinelles Lernen, neuronale Netzwerke, natürliche Sprachverarbeitung, Robotik und andere. Maschinelles Lernen ist ein zentraler Bestandteil von KI und ermöglicht es Maschinen, aus Erfahrungen zu lernen und sich selbst zu verbessern, ohne explizit programmiert zu werden.

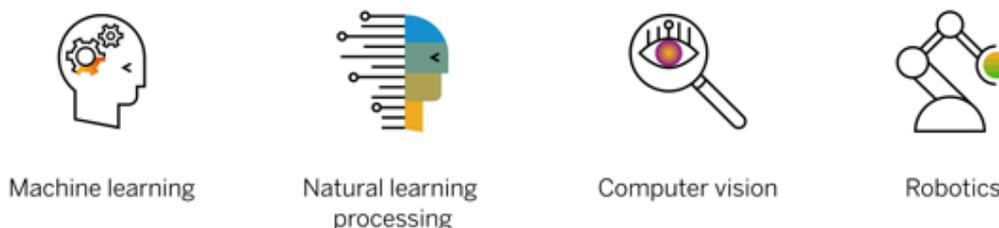


Abbildung 67: KI-Technologien

vgl.⁷² (SAP)

KI-Technologien sind entscheidend für die Erschließung des Potenzials künstlicher Intelligenz. Sie fungieren als Werkzeuge, die die Ideen des menschlichen Gehirns in praxisrelevante Erkenntnisse umsetzen.

Maschinelles Lernen (*engl.* Machine Learning (ML)) ist ein Teil der KI, nutzt Algorithmen zum selbstständigen Lernen und Verbessern aus Erfahrungen. Unternehmen setzen es ein, komplexe Datenanalysen durchzuführen und präzise Vorhersagen zu treffen.

Verarbeitung natürlicher Sprache (*engl.* Natural Language Processing (NLP)) ermöglicht es Maschinen, geschriebene oder gesprochene Sprache zu erkennen und zu interpretieren, wie z.B. in Chatbots und digitalen Sprachassistenten wie Siri und Alexa genutzt.

Computer Vision ermöglicht Computern, digitaler Bilder und Videos zu "sehen" und zu "verstehen". Anwendungen können komplexe Informationen auslesen und sogar durch Wände und um Ecken "durchscheuen", wie z.B. selbstfahrende Autos.

Robotik entwickelt die Automatisierung weiter durch die Integration der KI in der Hard- und Software, wie z.B. mit dem Einsatz von IoT-Sensoren, um eine bessere Abdeckung und Effizienz von Robotersteuerungsaufgaben in verschiedenen Bereichen zu steigern.

Diese Technologien der **künstlichen Intelligenz** (KI) verbessern nicht nur die Effizienz, sondern bieten auch innovative Lösungen für komplexe Herausforderungen in verschiedenen Branchen. Die nahtlose Integration künstlicher Intelligenz ebnet den Weg für ein zukunftsorientiertes Technologiezeitalter. Anwendungen der künstlichen Intelligenz werden häufig im Gesundheitswesen, im Finanzwesen, im Transportwesen, in der Unterhaltung, in der Automatisierung und in anderen Bereichen eingesetzt, die ethische und soziale Implikationen erfordern.⁷³

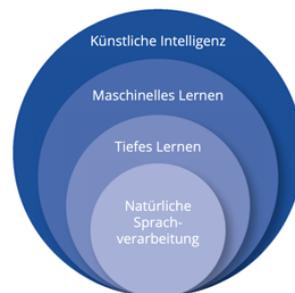


Abbildung 68: Leistungsbestandteile der Künstlichen Intelligenz

3.1.2.1. Machine Learning (ML)

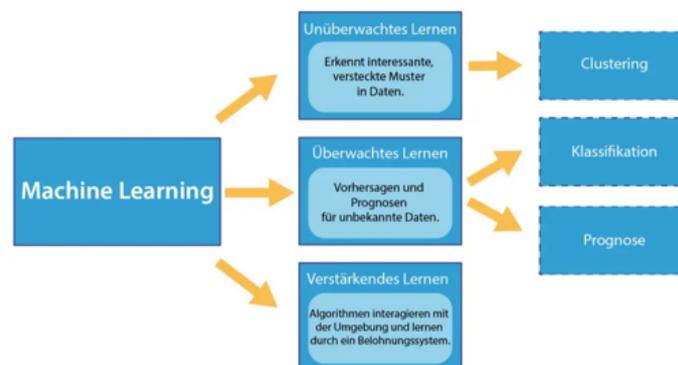


Abbildung 69: Arten von Machine Learning Algorithmen

Im folgendem wird das Machine Learning beschrieben:^{74 75}

Unüberwachtes Lernen (*engl.* unsupervised learning) ist ein Algorithmus, eingeschlossene Strukturen und Muster in den Daten eigenständig erkennen und zu lernen, wie z.B. ähnlichkeitsbasierte Clusteranalyse, welche ähnliche Objekte zu Gruppen zusammenfasst; Dimen-

vgl.⁷³ (BSI)

vgl.⁷⁴ (Wuttke)

vgl.⁷⁵ (Trabold, 2021)

sionsreduktion, dreidimensionalen Objekte, die zweidimensionale Abbildung liefern; Daten mit hochdimensionalen Objekten, die in niedrigere Dimensionen projizieren, um die Komplexität zu reduzieren. Hierbei wird der Algorithmus eigenständige interessante, verborgene Gruppen und Muster in den Daten ermittelt und von Data Scientists bewertet Ergebnisse werden durch verschiedene "weiche" Faktoren analysiert, um eweilige Anwendungen im Geschäftskontext zu bewerten. Hierbei können Anwendungen sein, wie z.B. Visualisierung großer Datenmengen, Clusteranalysen, Regelnextraktion und Merkmalerstellungen für ML sowie bei der Erkundung und Analyse von Daten, ohne dass explizite Labels oder Zielvariablen vorgegeben sind.

Model trainiert ohne Zielvariable und findet eigenständig Muster und Zusammenhänge in den Daten.

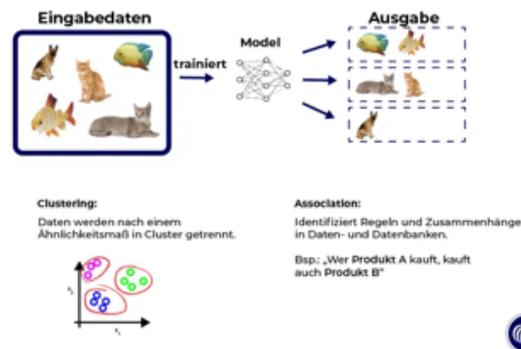


Abbildung 70: Unsupervised Learning (Unüberwachtes Lernen) ist eine Art von Machine Learning, als eigenständiges Muster und Zusammenhänge in den Daten findet

Überwachtes Lernen (engl. Supervised Machine Learning) beruht auf bekannten Daten, um Muster und Zusammenhänge zu identifizieren. Dabei werden Paradigma trainiert in den Algorithmus anhand eines definierten Datensatzes mit Trainingsdaten mit Zielvariable agieren, die korrekt vorhergesagt werden sollen, dabei entsprechen diese Zielvariable eine Klasse, wie z.B. Anwendungen wie, die Vorhersage der Kundenkündigung oder einen numerischen Wert, wie z.B. den Umsatz für den nächsten Monat, Vorhersage des Stromverbrauchs über einen bestimmten Zeitraum, die Risikobewertung von Investitionen, die Berechnung von Ausfallwahrscheinlichkeiten im Maschinenpark sowie die Prognose des Kundenwerts für die nächsten 12 Monate. Überwachtes Lernen umfasst das Trainieren von Klassifikations- oder Regressionsmodellen, die in verschiedenen Anwendungsbereichen universell einsetzbar sind. Dabei dienen **Klassifikationsmodelle** der Beantwortung von Fragen mit einer festen Anzahl von Antwortmöglichkeiten, während **Regressionsmodelle** treffen numerische Vorhersagen, wie z.B. die Vorhersage von Temperaturen. Beide Lernansätze spielen eine bedeutende Rolle in der Datenanalyse und insbesondere im Marketing, wo überwachtes Lernen oft zur Klassifikation von Kundendaten eingesetzt wird.

Model trainiert anhand von bekannten Daten und Beispielen (bspw. Hund vs. Katze). Es gibt eine klare Zielvariable, die vorhergesagt wird.

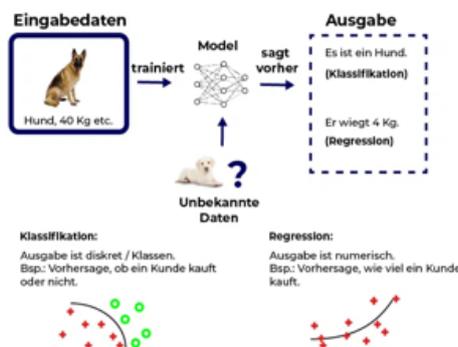


Abbildung 71: Überwachtes maschinelles Lernen trainiert Muster und Zusammenhänge anhand von Daten mit einer Zielvariable

Teilüberwachtes Lernen (engl. Semi-supervised Machine Learning) vereint Elemente des überwachten und unüberwachten Lernens. Dabei werden Trainingsdaten nur eine begrenzten Menge an Daten mit bekannten Zielvariablen als auch mit einer großen Menge

unbekannter Daten genutzt, um eine effiziente Datennutzung zu erstellen, da die Beschaffung bekannter Trainingsdaten sehr aufwendig und kostenintensiv ist und oft manuell durch Menschen erstellt wird. Z.B. sind manuelle Labeling von Bildern mit künstlichen neuronalen Netzen zur Klassifikation trainiert und auf den Rest der Daten angewendet werden, um die Trainingsdaten für die unbekanntenen Daten schneller und effizienter zu generieren. Es können Feedbacks als Betrugsfälle erstellt werden, wenn es sich um Betrug handelt, während für die übrigen Daten die Wahrscheinlichkeit besteht, dass es sich um einen geringeren Betrug handelt. Diese erstellten Ansätze lösen komplexe Probleme wie Betrugserkennung oder Bilderkennung und ermöglichen eine effektive Datennutzung, sowohl mit als auch ohne bekannten Zielvariablen.

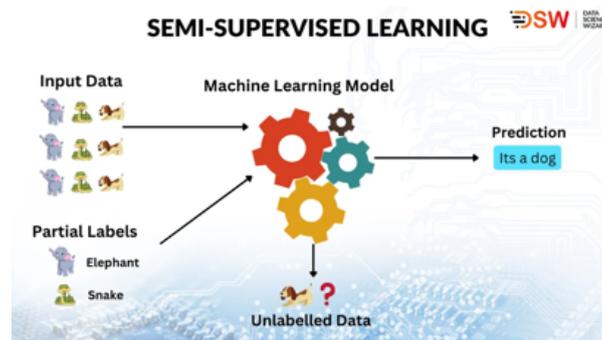


Abbildung 72: Semi-überwachten Lernen

Verstärkendes Lernen (*engl.* Reinforcement Learning) hier agieren die Algorithmen aktiv in ihrer Umgebung und zeichnen sich aus durch eine Kostenfunktion oder ein Belohnungssystembewertung, um eigenständige Strategie zur Problemlösung zu erlernen und dabei die Belohnung zu maximieren. Dabei erhält der Algorithmus positive und negative Feedbackrückmeldungen durch eine Kostenfunktion und erlernt, welche Handlungen in bestimmten Situationen angemessen sind und optimiert auf diese Weise die Belohnungsfunktion durch verstärkendes Lernen sind keine expliziten Datensätze erforderlich, da der Algorithmus in einer simulierten Umgebung interaktiv eine eigene Strategie entwickelt, wie z.B. Google DeepMind mit KI eine eigenständige erlernt. Diese eigenständigen Exploration macht das verstärkende Lernen besonders vielversprechend für komplexe Problemlösungen wie autonomes Fahren, autonome Robotik und die Entwicklung allgemeiner KI sowie durch Training erhält der Algorithmus oft erst nach vielen Schritten Feedback, um am Ende eines Trainings diese Erfahrungen zu nutzen, um eine effektive Strategie zu entwickeln.



Abbildung 73: einfaches Beispiel von verstärkendem Lernen durch Belohnungen

Aktives Lernen (*engl.* Active Learning) denen es unmöglich oder sehr kostspielig ist, die richtigen Antworten für alle Datenpunkte zu erhalten, wie z.B. sind Empfehlungssysteme. Dabei wird das Modell gezielt auswählen, welche Datenpunkte es für das Training verwenden möchte von denen das Modell am meisten lernen kann durch die Auswahl von Datenpunkten, bei denen das Modell unsicher ist oder bei denen es erwartet, dass sie einen hohen Informationsgewinn liefern. Hier wird die Trainingsprozesseffizienz erheblich verbessert, da das Modell mit weniger Trainingsdaten ähnliche oder sogar bessere Leistungen erzielen kann als bei der Verwendung aller verfügbaren Daten. Anwendungen finden in den Bereichen, bei der Beschaffung von Trainingsdaten, die teuer oder zeitaufwändig sind, wie z.B.

Empfehlungssysteme, Textklassifikation, Bilderkennung und medizinische Diagnose. Durch die gezielte Trainingsdatenauswahl können Modelle schneller und präziser trainiert werden, dass zu einer besseren Entscheidungen und Ergebnissen führt.⁷⁶

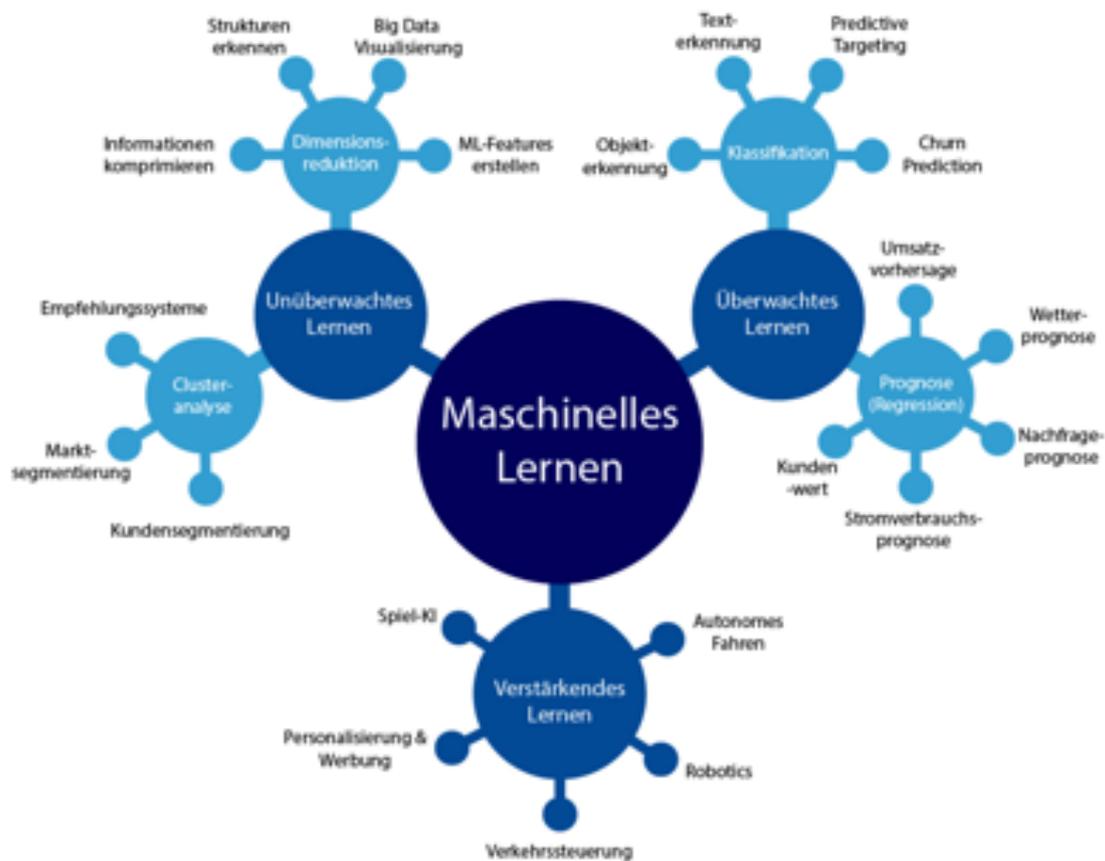


Abbildung 74: Maschinelles Lernen im Überblick: Anwendungsbeispiele nach Arten

3.1.2.1.2. ML mit Neural Network

Nach IBM-Wissensstand wird der nachfolgende Abschnitt über neuronale Netze anlehend beschrieben:⁷⁷

Künstliche neuronale Netze (KNN), auch als simulierte neuronale Netze (SNN) bezeichnet, sind ein zentraler Bestandteil des Deep Learning und gehören zur Disziplin des maschinellen Lernens (ML). Ihre Struktur und Benennung sind dem menschlichen Gehirn nachempfunden, wobei sie die Funktionsweise biologischer Neuronen imitieren, indem sie Signale zwischen einzelnen Knoten weiterleiten. Ein künstliches neuronales Netz besteht aus verschiedenen Schichten von Knoten: einer Eingabeschicht, einer oder mehreren versteckten Schichten und einer Ausgabeschicht. Jeder dieser Knoten – auch als künstliches Neuron bezeichnet – ist mit anderen Knoten verbunden und besitzt individuelle Gewichtungen sowie einen Schwellenwert. Wenn die Ausgabe eines Knotens den Schwellenwert überschreitet, wird er aktiviert und leitet die Informationen an die nächste Schicht weiter. Andernfalls erfolgt keine Datenweitergabe. Die drei wesentlichen Komponenten eines neuronalen Netzes sind die Trainingsdaten, das Modell und der Lernalgorithmus, der das Modell darauf trainiert, Muster in den Daten zu erkennen. Die Inspiration für künstliche neuronale Netze stammt aus der Neurowissenschaft, in der neuronale Verbindungen als Netzwerke betrachtet werden, die bestimmte Funktionen im Nervensystem ausführen. Künstliche neuronale Netze orientieren sich an diesen Organisationsprinzipien biologischer neuronaler Netze und setzen diese mithilfe mathematischer Modelle um, um komplexe Aufgaben zu bewältigen.

vgl.⁷⁶ (Wuttke, 2024)
vgl.⁷⁷ (IBM)

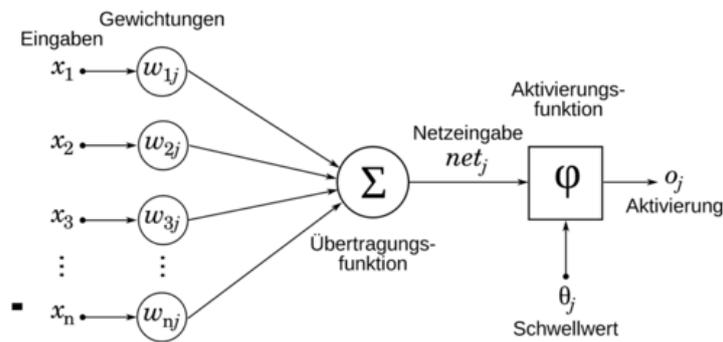


Abbildung 75: neuronales Netzwerk

Sowohl biologische als auch künstliche Neuronen empfangen, verarbeiten und leiten Informationen weiter. Ein wesentliches Merkmal neuronaler Netze ist ihre nicht-lineare Informationsverarbeitung, die durch die Verknüpfung von Neuronen und spezielle Verarbeitungsfunktionen ermöglicht wird. Dies erlaubt die Darstellung komplexer Abhängigkeiten der ursprünglichen Informationen. Von großer Bedeutung ist dabei, dass neuronale Netze diese Abhängigkeiten autonom erlernen können, basierend auf Erfahrungsdaten, den sogenannten Trainingsdaten, mit denen sie gefüttert werden. Ein neuronales Netzwerk besteht aus mehreren Schichten, wobei die Eingabeschicht die Rohdaten enthält. Jede weitere Schicht erhält den Output der vorherigen Schicht und nicht mehr die ursprünglichen Daten. Das KI-System lernt von jedem Übergang zu einer weiteren Schicht dazu. Die letzte Schicht, die Ausgangsschicht, erzeugt die Ergebnisse. Das System trainiert sich somit selbst, indem es die ursprünglichen Daten und Regeln verwendet und dieser Prozess wird als maschinelles Lernen bezeichnet.

Die Vorstellung eines jeden Knotens als eigenes Modell einer linearen Regression ist grundlegend für das Verständnis neuronaler Netzwerke. Ein solcher Knoten setzt sich aus Eingabedaten, Gewichtungen, einer Verzerrung (Bias) oder einem Schwellenwert sowie einer Ausgabe zusammen. Die mathematische Formel für diesen Prozess lautet:

$$\sum w_i x_i + b \text{ Bias} = w_1 x_1 + w_2 x_2 + w_3 x_3 + \text{Bias}$$

$$\text{output} = f(x) = 1 \text{ if } \sum w_1 x_1 + b \geq 0; 0 \text{ if } \sum w_1 x_1 + b < 0$$

Die Ausgabe eines Knotens in einem künstlichen neuronalen Netz wird durch eine Aktivierungsfunktion bestimmt, die entscheidet, ob der Knoten aktiviert wird und Daten an die nächste Schicht weiterleitet. Dadurch wird die Ausgabe eines Knotens als Eingabe für den nächsten verwendet, was die Architektur des Netzwerks als Feedforward-Netz charakterisiert. Ein praxisnahes Beispiel ist der Einsatz eines neuronalen Netzwerks in einem Netzwerküberwachungssystem eines Unternehmens zur Erkennung verdächtiger Aktivitäten. In diesem Szenario wird das neuronale Netz dazu genutzt, Muster in den Netzwerkdaten zu identifizieren und potenzielle Sicherheitsbedrohungen zu erkennen. Die Eingabedaten für das neuronale Netzwerk könnten verschiedene Netzwerkparameter und Metriken umfassen, wie etwa die Klassifikation eines Vorfalls: Sicherheitsbedrohung vorhanden (**Ja = 1, Nein = 0**), Netzwerk online (**Ja = 1, Nein = 0**), oder Angriff erfolgt (**Ja = 1, Nein = 0**). Das neuronale Netz gibt als Vorhersageergebnis (\hat{y}) aus, ob eine Bedrohung vorliegt und es einfließen in dieser Entscheidungsfindung, z.B. Anzahl fehlgeschlagener Anmeldeversuche, Anzahl von Dateiübertragungen mit verdächtigen IP-Adressen, Anomalien im Datenverkehrsvolumen sowie Häufigkeit von DNS-Anfragen an nicht autorisierte Domänen, ein. Hierbei könnten potenzielle Cyberangriffe darauf hingewiesen werden und fundierten Entscheidungen über das Vorliegen eines Sicherheitsvorfalls schneller getroffen werden.

Angenommen, die Eingabedaten lauten wie folgt:

1. $X_1 = 1$ (Anzahl fehlgeschlagener Anmeldeversuche)
2. $X_2 = 0$ (Anzahl von Dateiübertragungen auf verdächtige IP-Adressen)
3. $X_3 = 1$ (ein Cyber-Angriff)

Diese Eingabedaten werden in das neuronale Netzwerk eingespeist, das aus mehreren Schichten von Neuronen besteht. Jede Schicht verarbeitet die Eingabedaten und leitet sie an die nächste Schicht weiter, wobei komplexe Muster erkannt werden. Die Gewichtungen in den Neuronen des Netzwerks werden während des Trainingsprozesses angepasst, um das Netzwerk zu optimieren und die Genauigkeit bei der Erkennung von Sicherheitsbedrohungen zu verbessern.

- W1 = 5, Anzahl fehlgeschlagener Anmeldeversuche sind sehr hoch
- W2 = 2, Anzahl von Dateiübertragungen von unterschiedlichen verdächtige IP-Adressen
- W3 = 4, Anomalien im Datenverkehrsvolumen weisen auf ein Cyber-Angriff hin

Die Gewichtungen werden zugewiesen, die den Einfluss jeder Variable auf die Entscheidung bestimmen. Größere Gewichtungen bedeuten einen stärkeren Einfluss. Abschließend wird ein Schwellenwert von 3 angenommen unter Berücksichtigung dieser Eingaben und Gewichtungen ergibt sich die Berechnung für das vorhergesagte Ergebnis:

$$\hat{y} = (1 \cdot 5) + (0 \cdot 2) + (1 \cdot 4) - 3 = 6$$

Die Aktivierungsfunktion würde hier eine Ausgabe von **1** ergeben, da **6 größer als 0** ist. Nach dem Training kann das neuronale Netzwerk in Echtzeit auf die eingehenden Netzwerkdaten angewendet werden. Wenn das Netzwerk eine Anomalie oder verdächtige Aktivität erkennt, kann es eine Warnung ausgeben oder automatisierte Maßnahmen zur Eindämmung der Bedrohung ergreifen. Durch die kontinuierliche Anpassung und Optimierung des neuronalen Netzwerks kann das Unternehmen seine Sicherheitsüberwachung verbessern und proaktiv auf potenzielle Bedrohungen reagieren. Obwohl Perzeptrone genutzt wurden, um mathematische Zusammenhänge zu veranschaulichen, verwenden neuronale Netze Sigmoidneuronen, die Werte **zwischen 0 und 1** haben. Durch den Informationsfluss von einem Knoten zum nächsten reduziert sich die Auswirkung einzelner Variablenänderungen auf die Netzwerkausgabe.

Bei praktischen Anwendungen wie Bilderkennung oder Klassifizierung ist das Training des Modells mit überwachtem Lernen und markierten Datensätzen üblich. Die Genauigkeit des Modells wird durch eine Kostenfunktion wie den mittleren quadratischen Fehler bewertet. Durch Gradientenabstieg und bestärkendes Lernen passt sich das Modell an, um die Kostenfunktion zu minimieren und so die Korrektheit für jede Beobachtung sicherzustellen.

$$MSE = \frac{1}{2} \sum_{i=1}^{129} (y_i - \hat{y}_i)^2$$

i steht für den Index der Stichprobe, \hat{y} hat für das vorhergesagte Ergebnis, *y* für den tatsächlichen Wert und *m* für die Anzahl der Stichproben.

MSE - mittlere quadratische Fehler

y_i - beobachteten Wert für den *i*-ten Datenpunkt mit **6**

\hat{y}_i - vorhergesagten Wert für den *i*-ten Datenpunkt mit **2**

\sum - summiert die quadrierten Unterschiede zwischen den beobachteten und vorhergesagten Werten über alle Datenpunkte von **i=1 bis i=129**

1/2 wird in einigen Fällen verwendet, um die Berechnungen zu vereinfachen, hat aber keinen Einfluss auf das Ergebnis.

$$MSE = \frac{1}{2} \sum_{i=1}^{129} (6 - 2)^2$$

- Differenz zwischen dem **vorhergesagten Wert (66)** und dem **beobachteten Wert (22)** beträgt **6-2=4**
- Differenz wird quadriert, **(4)²=16**
- Wert für jeden Datenpunkt konstant ist, bleibt er während der Summation unverändert

- Summe \sum summiert diese quadrierten Unterschiede über alle Datenpunkte von **i=1 bis i=129**

$$\sum_{i=1}^{129} (y_i - \hat{y}_i)^2 = 129 \times 16 = 2064$$

- **1/2** ist ein Skalierungsfaktor wird manchmal verwendet, hat aber keinen Einfluss auf das Ergebnis, da er nur zur Vereinfachung der Berechnungen dient
- Wert für jeden Datenpunkt konstant ist und wir **129 Datenpunkte** haben, beträgt das Ergebnis des MSE:

$$\text{MSE} = \frac{1}{129} \times 2064 = 16$$

Das Hauptziel eines Modells besteht darin, die Kostenfunktion zu minimieren, sodass es optimal an die vorliegenden Daten angepasst wird. Im Anpassungsprozess werden die Gewichtungen und Verzerrungen des Modells fortlaufend angepasst, wobei die Kostenfunktion eine zentrale Rolle spielt. Durch den Einsatz von Reinforcement Learning und Gradientenabstieg ermittelt das Modell schrittweise die Richtung, in der die Fehler reduziert werden können, um den Konvergenzpunkt oder zumindest ein lokales Minimum zu erreichen. Mit jeder Trainingsinstanz werden die Modellparameter weiter optimiert, um das Minimum der Kostenfunktion anzunähern. Dabei ist eine umfassende Datenerhebung entscheidend: Die Trainingsdaten sollten vielfältig und repräsentativ sein, um eine robuste und zuverlässige Anpassung des Modells zu gewährleisten.⁷⁸

wichtige neuronale Netzwerk-Arten sind:

Neuronale Netzwerke sind im Bereich des ML, der verschiedene Architekturen wie flache, tiefe und rekurrente Modelle umfasst.

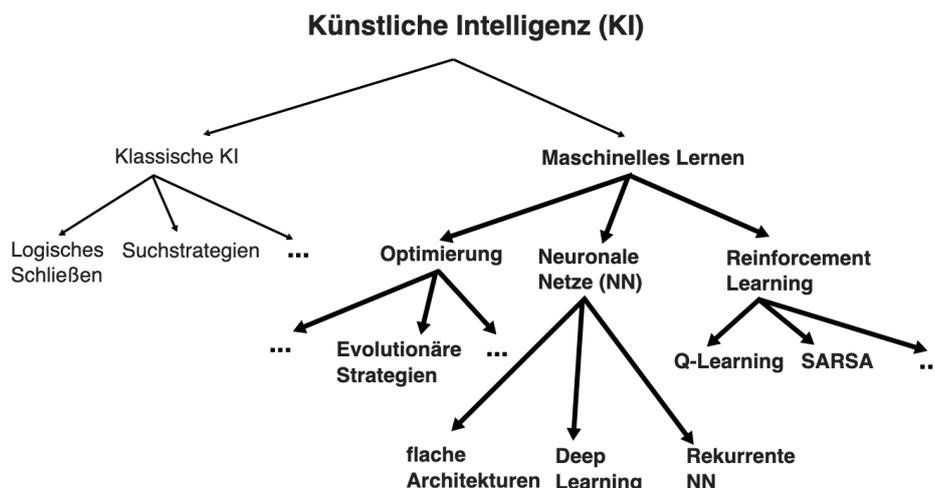


Abbildung 76: Einordnung neuronale Netz-Arten

3.1.2.1.3. Arten von Neural Network

Perzeptron-Neural Network ist das einfachste neuronale Netzwerk und nimmt Eingabeparameter entgegen, summiert diese auf, wendet eine Aktivierungsfunktion an und gibt das Ergebnis an die Ausgabeschicht weiter, als **binär**, also entweder **0 oder 1**, ähnlich einer **Ja-oder Nein-Entscheidung**. Die finale Entscheidung wird durch Vergleich des Aktivierungswerts mit einem Schwellwert getroffen. Überschreitet der Aktivierungswert den Schwellwert, wird dem **Ergebnis eine 1** zugewiesen; ansonsten wird eine **0 zugewiesen**. Basierend auf diesem Prinzip wurden weitere neuronale Netzwerke und Aktivierungsfunktionen entwickelt, die auch mehrere Ausgaben mit **Werten zwischen 0 und 1** ermöglichen, wie z.B. die **Sigmoid-Funktion**, die bei solchen Anwendungen als Aktivierungsfunktion Verwendung findet. **Multilayer-Perzeptron-Neural Networks (MLP)** steigert die Komplexität und haben eine verborgene Schicht in Ihren Perzeptron-Netzwerken.⁷⁹

vgl.⁷⁸ (IBM)

vgl.⁷⁹ (CLOUDFLARE)

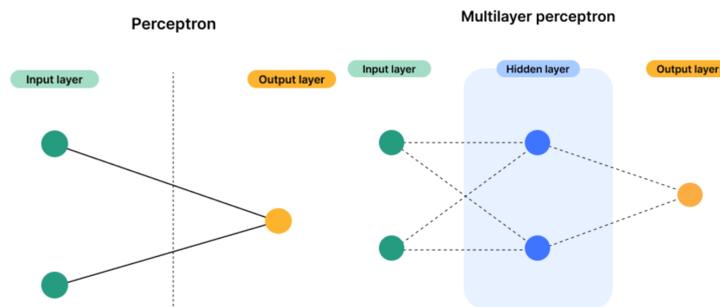


Abbildung 77: einfache und Multilayer neuronale Perceptron

Der Einsatz in SIEM-Anwendungen können sein, wie z.B. **Anomalieerkennung** sind durch Training auf historischen Daten fixiert und lernen diese Netzwerke normale Systemverhalten und identifizieren Abweichungen als potenzielle Bedrohungen, die ideal sind zur Erkennung von Zero-Day-Angriffen; **Automatisierte Bedrohungsanalyse** ermöglichen es, die automatische Klassifizierung und Priorisierung sicherheitsrelevanter Ereignisse und das Sicherheitsteams zu entlasten und auf die kritischsten Vorfälle zu fokussieren; **Optimierung der Alarmgenauigkeit** werden durch Training präzises Fehlalarme reduziert und verbessert die Unterscheidung zwischen echten Bedrohungen und harmlosen Anomalien.

Feedforward-Neural Network (FFN) sind die Schichten nur mit der unmittelbar folgenden Schicht verbunden und es gibt keine Rückkopplungsschleifen. Der Trainingsprozess eines verläuft in der Regel so, dass alle Knoten sind miteinander verbunden sind und die Aktivierung erfolgt von der Eingangsschicht zur Ausgangsschicht und es gibt mindestens eine bzw. viele Zwischenschicht (Layer) zwischen der Eingangs- und der Ausgangsschicht.⁸⁰

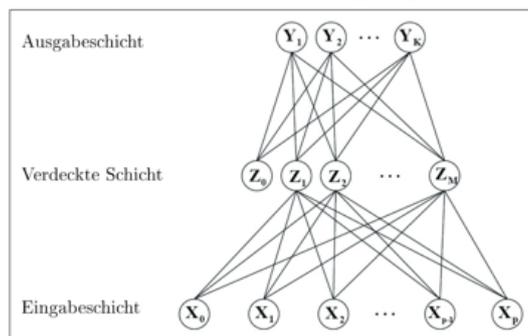


Abbildung 78: Netzwerkdiagramm eines Feedforward-Netzes

Der Einsatz in SIEM-Anwendungen können sein: **Effektive Mustererkennung** lernen typische Verhaltensmuster anhand historischer Sicherheitsdaten und erkennen Abweichungen als potenzielle Bedrohungen, ideal für die Erkennung bekannter und neuer Angriffe, einschließlich Zero-Day-Exploits; **Erhöhte Automatisierung** beschleunigen die Sicherheitsanalyse durch automatisierte Klassifizierung und Priorisierung von Vorfällen, wodurch sich Sicherheitsteams auf die kritischsten Bedrohungen fokussieren können; **Verbesserte Alarmgenauigkeit** reduzieren Fehlalarme durch präzise Unterscheidung zwischen tatsächlichen Bedrohungen und harmlosen Anomalien.

Convolutional Neural Networks (CNN), auch als faltende neuronale Netzwerke bekannt, sind spezialisierte künstliche neuronale Netzwerke, die mit 2D- oder 3D-Eingabedaten arbeiten und häufig für die Objekterkennung in Bildern verwendet werden. Der wesentliche Unterschied zu herkömmlichen neuronalen Netzwerken liegt in der spezifischen Architektur von CNNs, die durch die sogenannte Faltung („**Convolution**“) gekennzeichnet ist. Die verborgenen Schichten eines CNN bestehen aus einer Abfolge von Faltungs- und Pooling-Operationen und können während der Faltung ein sogenannter Kernel über die Eingabedaten verschieben, wobei eine Faltungsoperation durchgeführt wird, die einer Multiplikation ähnelt. Dabei werden die Neuronen entsprechend aktualisiert und anschließend wird eine Pooling-Schicht eingeführt, um die Ergebnisse zu vereinfachen und nur die relevanten

vgl.⁸⁰ (Wallner, 2007)

Informationen zu extrahieren. Dieser Prozess führt zu einer allmählichen Verkleinerung der 2D- oder 3D-Eingabedaten und am Ende dieses Prozesses entsteht in der Ausgabeschicht ein Vektor, der als „**fully connected layer**“ bezeichnet wird. Dieser Vektor ist besonders in Klassifikationsaufgaben von großer Bedeutung, da er Wahrscheinlichkeiten für verschiedene Klassen berechnet und so viele Neuronen enthält, wie es Klassen gibt, um eine präzise Zuordnung vorzunehmen.⁸¹

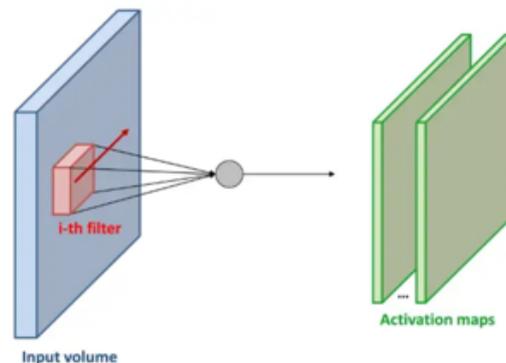


Abbildung 79: Faltung in Convolutional Neural Networks

Der Einsatz in SIEM-Anwendungen können sein: **Anomaliedetektion** zu erkennen von Muster in sicherheitsrelevanten Daten wie Netzwerkverkehr und Logdateien und zu lernen, zwischen normalem und abweichendem Verhalten zu unterscheiden, um Sicherheitsvorfälle frühzeitig zu erkennen; **Einsatz in Intrusion Detection Systems (IDS)** finden in Netzwerkaktivitäten und Protokollanalysen statt, um verdächtige Verhaltensweisen oder Angriffe zu identifizieren und die Erkennungseffizienz zu erhöhen; **Phishing-Erkennung** dort werden verdächtige URLs, E-Mails oder Webseiten-Inhalte analysiert und typische Phishing-Muster von legitimen Inhalten unterscheiden; **Analyse von Malware** werden Merkmalen wie ausführbare Dateien und Netzwerkverhalten auf bestehenden Malware-Daten trainiert; **Verhaltensanalyse von Nutzern** überwachen Benutzerverhalten und identifizieren untypische Aktivitäten, die auf Sicherheitsprobleme hindeuten könnten; **Erkennung von DDoS-Angriffen** helfen bei der Erkennung von DDoS-Angriffen, um Anomalien im Netzwerkverkehr zu erkennen und ungewöhnliche Verkehrsmuster zu identifizieren.⁸²

Recurrent Neural Networks (RNN) erweitern klassische neuronale Netzwerke durch die Einführung wiederkehrender Zellen, die dem Netzwerk die Fähigkeit verleihen, eine Art Gedächtnis zu entwickeln. Diese Architektur ermöglicht es dem Netzwerk, Informationen nicht nur in eine Richtung zu verarbeiten, sondern auch rückwärts, sodass die Ausgabe bestimmter Knoten die Eingabe früherer Knoten beeinflussen kann. Dabei erhält jede versteckte Zelle ihre eigene Ausgabe mit einer festgelegten Verzögerung – gespeichert für eine oder mehrere Iterationen –, was RNNs grundsätzlich von klassischen Feedforward-Netzwerken unterscheidet. Trotz verschiedener Variationen, wie etwa der Übergabe von Zuständen an Eingabeknoten oder der Nutzung variabler Verzögerungen, bleibt das Kernprinzip bestehen. RNNs werden dann verwendet, wenn der Kontext entscheidend ist, da vergangene Iterationen wesentlich zur Entscheidungsfindung in aktuellen Iterationen beitragen können. Diese Eigenschaft macht sie besonders geeignet für Aufgaben wie die Verarbeitung von Sequenzdaten, bei denen zeitliche Abhängigkeiten eine zentrale Rolle spielen. Der Nachteil der RNNs ist über die Zeit instabil, was zu Schwierigkeiten bei der Modellierung langfristiger Abhängigkeiten führen kann, deshalb werden heutzutage häufig **Long Short-Term Memory Units (LSTM)** verwendet, um die Stabilität von RNN zu verbessern. LSTM sind spezielle Zellen, die dazu dienen, langfristige Abhängigkeiten besser zu erfassen und zu modellieren. Ein häufiges Anwendungsbeispiel für RNN und LSTM ist die Textverarbeitung, bei der die Bedeutung eines Wortes oft stark von vorherigen Wörtern oder Sätzen abhängt. Aber auch die Verarbeitung von Videodaten findet hier eine Anwendung, wie etwa bei der Objekter-

vgl.⁸¹ (Wuttke, 2024)

vgl.⁸² (Wallner, 2027)

kennung und -verfolgung im Bereich des autonomen Fahrens, wo Objekte innerhalb von Bildsequenzen erkannt und über die Zeit hinweg verfolgt werden.⁸³

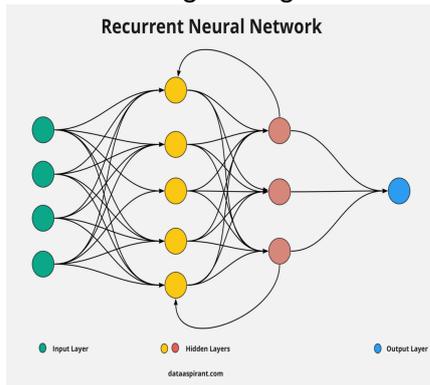


Abbildung 80: Aufbau eines Recurrent Neural Networks und Long Short-Term Memory Units

Der Einsatz in SIEM-Anwendungen kann sein: **Analyse von Zeitreihen** werden verarbeitet durch zeitlich strukturierte Daten wie Netzwerkprotokolle und Logdateien, um zeitliche Abhängigkeiten zu erkennen und mögliche Sicherheitsvorfälle oder abweichendes Verhalten zu identifizieren; **Anomaliedetektion** werden Abweichungen in Datenströmen identifiziert, indem diese typische Verhaltensmuster lernen und dadurch können Anomalien aufgedeckt werden, die auf Sicherheitsbedrohungen hinweisen, und frühzeitig Alarm schlagen; **Erkennung von Insider-Bedrohungen** überwachen langfristig Benutzerverhalten, erkennen subtile Abweichungen von normalen Mustern und identifizieren so potenzielle Insider-Bedrohungen frühzeitig; **Korrelation von Protokolldaten** werden in der Log-Analyse erkannt und mit RNN werden zusammenhängende Ereignisse bewertet, die auf koordinierte Angriffe oder Sicherheitsprobleme hindeuten, und helfen, komplexe Bedrohungen besser zu verstehen; **Vorhersage von Sicherheitsvorfällen** nutzen historische Daten, um Muster zu erkennen, die auf zukünftige Sicherheitsvorfälle hinweisen, und ermöglichen so proaktive Maßnahmen zur Risikominderung.⁸⁴

Modulare-Neural Networks (MNNs) kombinieren mehrere spezialisierte neuronale Netze, um gemeinsam eine konsolidierte Ausgabe zu erzeugen. Jedes Modul bearbeitet spezifische Teilaufgaben, die anschließend zu einer Gesamtlösung zusammengeführt werden. Diese Architektur ist besonders nützlich bei Problemen, die unterschiedliche Verarbeitungstechniken erfordern, wie die Kombination von Bild- und Textverarbeitung. MNNs bieten durch ihre Modularität eine hohe Flexibilität, Spezialisierung und verbesserte Generalisierungsfähigkeit, während das Training und die Optimierung erleichtert werden, da jedes Modul separat angepasst werden kann.⁸⁵

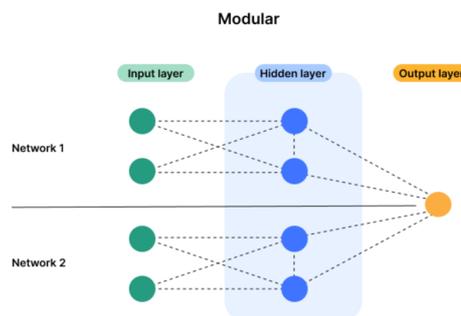


Abbildung 81: Modulare neuronale Netzwerke (MNNs)

Der Einsatz in SIEM-Anwendungen kann sein: **Analyse von Zeitreihen**, die zeitlich geordnete Daten verarbeiten wie Netzwerkverkehr und Protokolle, Muster über die Zeit erkennen und dadurch Anomalien oder Sicherheitsvorfälle aufdecken; **Erkennung von Anomalien** durch identifizieren ungewöhnliche Muster in sequentiellen Daten und sind

vgl.⁸³ (Grellmann, 2021)

vgl.⁸⁴ (Wuttke)

vgl.⁸⁵ (CLOUDFLARE)

dadurch besonders effektiv bei der Erkennung von Anomalien, die auf Bedrohungen hindeuten; **Identifikation von Insider-Bedrohungen** werden langfristige Benutzerverhalten überwacht, wiederkehrende oder abweichende Muster erkennt und mögliche Insider-Bedrohungen identifiziert; **Korrelation von Protokolldaten** werden Protokolldaten analysiert, um zusammenhängende Ereignisse zu erkennen, die auf Sicherheitsprobleme oder Angriffe hindeuten; **Vorhersage von Sicherheitsvorfällen** werden historische Daten analysiert, um Muster zu identifizieren, die auf zukünftige Sicherheitsvorfälle hinweisen, und ermöglichen so präventive Maßnahmen.

Radialen Basisfunktionen-Neural Network (RBF) verwenden spezielle radialsymmetrische Aktivierungsfunktionen, deren Werte mit zunehmender Distanz zum Mittelpunkt abnehmen. Eine häufige RBF ist die Gaußsche Funktion, die eine glockenförmige Kurve darstellt wird. In RBF-Netzwerken misst jeder Knoten im versteckten Layer den Abstand zwischen einem Eingangsvektor und einem prototypischen Muster und je näher der Eingangsvektor am Prototyp liegt, desto höher die Aktivierung. Diese Netzwerke eignen sich besonders für die Modellierung nichtlinearer Zusammenhänge und werden häufig in der Klassifikation, Regression und Funktionennäherung eingesetzt, da sie schnelle Konvergenz und gute Ergebnisse auch bei kleineren Datensätzen bieten.⁸⁶

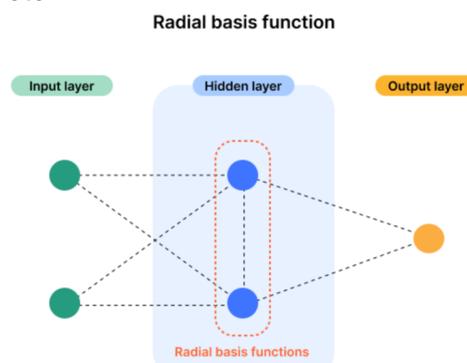


Abbildung 82: Radialen Basisfunktionen-Neuronale Netzwerke

Der Einsatz in SIEM-Anwendungen können sein: **Anomaliedetektion** es werden präzise Anomalien im Netzwerkverkehr und in Logdaten erkannt, indem Datenpunkte in einem hochdimensionalen Raum abbildet werden und Abweichungen von normalen Mustern frühzeitig identifiziert werden; **Mustererkennung** werden durch komplexe, nichtlineare Muster modelliert und sind besonders effektiv bei der Erkennung spezifischer Angriffsmuster eingesetzt, einschließlich neuer oder unbekannter Bedrohungen, die traditionelle Methoden übersehen könnten; **Intrusion Detection Systems (IDS)** helfen RBFs, Netzwerkaktivitäten zu klassifizieren und schädliche von normalen Aktivitäten zu unterscheiden, was die Erkennung und Abwehr von Angriffen verbessert; **Verhaltensanalyse** überwachen kontinuierlich das Verhalten von Benutzern und Systemen, erkennen ungewöhnliche Verhaltensänderungen und sind besonders nützlich zur Identifizierung von Insider-Bedrohungen; **Korrelationsanalyse** unterstützt die Korrelation von sicherheitsrelevanten Ereignissen in SIEM-Systemen, um komplexe Angriffsmuster zu verstehen und rechtzeitig zu erkennen.

Liquid State Machine-Neural Networks (LSM) sind eine Form von **Spiking Neural Networks (SNNs)**, die durch zufällig vernetzte Neuronen gekennzeichnet sind. Diese Netzwerke verarbeiten Eingangssignale dynamisch, wobei die zufällige Struktur flexible und komplexe Reaktionen ermöglicht. LSMs fungieren als Reservoir Computing-Systeme, die zufälligen Verbindungen im Netzwerk erzeugen einen dynamischen Zustand, der zeitliche und räumliche Informationen der Eingabe speichert. Nur der Auslesemechanismus wird trainiert, um aus diesem Reservoir die gewünschten Ausgaben zu erzeugen. LSMs sind besonders effektiv bei der Verarbeitung zeitabhängiger Daten, wie sie in Spracherkennung und Echtzeit-Signalverarbeitung vorkommen.⁸⁷

vgl.⁸⁶ (CLOUDFLARE)

vgl.⁸⁷ (CLOUDFLARE)

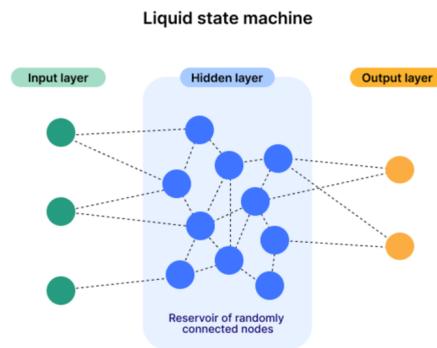


Abbildung 83: Liquid State Machine-Neuronale Netzwerke

Der Einsatz in SIEM-Anwendungen können sein: **Echtzeit-Anomalieerkennung** erkennen komplexe zeitliche Muster in Echtzeit, analysieren kontinuierlich Netzwerkverkehr und Systemprotokolle und identifizieren sofort Abweichungen, die auf Sicherheitsbedrohungen hinweisen könnten; **Verarbeitung zeitabhängiger Daten** analysieren zeitlich geordnete Daten, wie Netzwerkaktivitäten oder Benutzerinteraktionen, und erkennen Muster und Abhängigkeiten, um Angriffe zu identifizieren, die sich über die Zeit entwickeln; **Dynamische Bedrohungserkennung** passen sich flexibel an neue Bedrohungsszenarien an, indem sie kontinuierliche Veränderungen in Datenströmen verarbeiten, und sind somit in Echtzeit wertvoll für dynamische Sicherheitsumgebungen; **Erkennung von Insider-Bedrohungen** überwachen kontinuierlich Benutzer- und Systemverhalten, erkennen subtile, komplexe Verhaltensänderungen und identifizieren Insider-Bedrohungen frühzeitig; **Effiziente Korrelation von Sicherheitsereignissen** korrelieren sicherheitsrelevante Ereignisse über verschiedene Zeitpunkte hinweg, ermöglichen tiefgreifende Analysen und helfen, komplexe Bedrohungen frühzeitig zu erkennen und zu verstehen.

Residuale-Neuronale-Neural Network (ResNet) verbessern das Training tiefer Netzwerke durch das Konzept des **Identity Mapping** und die Ausgabe einer früheren Schicht direkt zur Ausgabe einer späteren Schicht addiert und somit eine Identitätsverbindung eingefügt wird. Diese Shortcut-Verbindungen ermöglichen das Überspringen von Schichten und sorgen dafür, dass ursprüngliche Informationen erhalten bleiben und den Gradientendurchfluss erleichtern. Dabei wird das Problem des **vanishing gradients** gelöst und ermöglicht das Training sehr tiefer Netzwerke, die in der Bildverarbeitung und anderen Anwendungen herausragende Ergebnisse erzielen. ResNet bildet die Grundlage für moderne Architekturen, die die Leistungsfähigkeit tiefer Netzwerke weiter steigern.⁸⁸

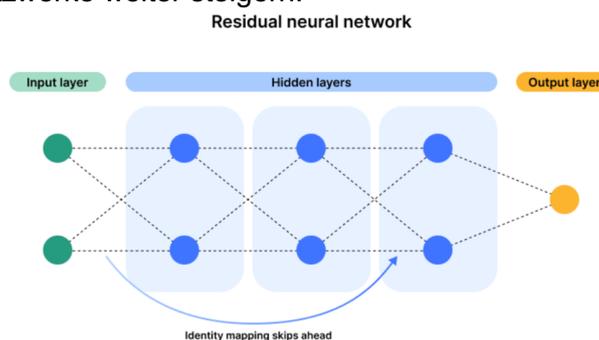


Abbildung 84: Residuale-Neuronale Netzwerke

Der Einsatz in SIEM-Anwendungen können sein: **Präzise Mustererkennung** erkennen effektiv komplexe Muster in großen Datensätzen, wie in SIEM-Systemen vorkommen, und identifizieren selbst feinste Anomalien, was die Erkennung von Sicherheitsvorfällen verbessert; **Effiziente Datenverarbeitung** können tiefere Netzwerke effizient trainieren und verarbeiten so große Mengen sicherheitsrelevanter Daten, was für die Analyse kontinuierlicher Datenströme in SIEM-Systemen entscheidend ist; **Genauere Anomaliedetektion** erfassen ResNets komplexe Anomalien präzise, was zu einer genaueren Erkennung von Sicherheits-

vgl.⁸⁸ (CLOUDFLARE)

risiken führt und Fehlalarme minimiert; **Robuste Merkmalsextraktion** mit komplexen Datensätzen relevante Merkmale zu extrahieren, was in SIEM-Systemen für die Bedrohungserkennung essentiell ist; **Hohe Flexibilität** lassen sich leicht an spezifische SIEM-Anforderungen anpassen und reagieren dynamisch auf neue Bedrohungen, was ihre Anpassungsfähigkeit in einer sich ständig verändernden Sicherheitslandschaft erhöht.

Generative Adversarial-Neural Network (GAN) besteht aus zwei konkurrierenden neuronalen Netzwerken mit einem Generator und einem Diskriminator. Der Generator erzeugt synthetische Daten, die echten Daten möglichst ähnlich sind, während der Diskriminator zwischen echten und generierten Daten unterscheidet. Durch das Zusammenspiel dieser Netzwerke verbessert sich der Generator kontinuierlich, bis er Daten produziert, die vom Diskriminator nicht mehr von echten Daten unterschieden werden können. GAN findet Verwendung besonders bei leistungsfähigen Anwendungen in der Erstellung realistischer Bilder, Kunstwerke, Text- und Musikgenerierung und haben sich in vielen kreativen und technischen Anwendungen bewährt. Trotz ihrer Stärke sind sie anspruchsvoll im Training und können anfällig für Instabilitäten und Modus-Kollaps sein, bei dem die generierte Vielfalt eingeschränkt ist.⁸⁹

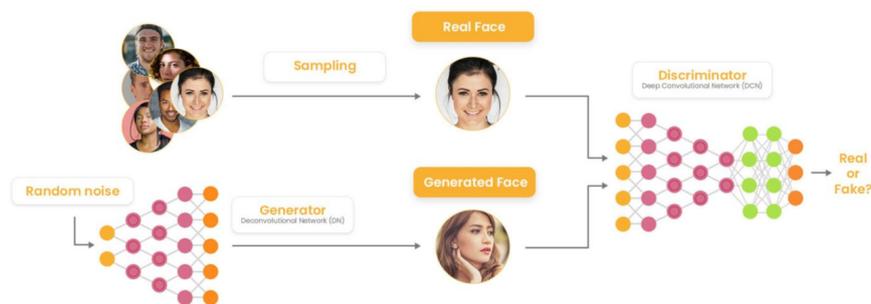


Abbildung 85: Generative Adversarial Networks

Der Einsätze in SIEM-Anwendungen können sein: **Präzise Anomalieerkennung** lernen typische Verhaltensmuster in Netzwerkverkehr und Logdaten und identifizieren Abweichungen, was zu einer genauen Erkennung von Anomalien und Sicherheitsbedrohungen führt, selbst bei subtilen Abweichungen; **Erzeugung synthetischer Daten** generieren realistische, synthetische Daten, die zur Schulung von Sicherheitsmodellen verwendet werden können, insbesondere wenn echte Bedrohungsdaten schwer verfügbar sind, um SIEM-Systeme robuster gegen neue Bedrohungen zu machen; **Verstärkung von Intrusion Detection Systems (IDS)** verbessern IDS, indem sie realistische, bösartige Muster erzeugen, die zur Schulung des Systems verwendet werden, wodurch die Erkennung und Reaktion auf Angriffe optimiert wird; **Modellierung von Benutzerverhalten** modellieren normale Verhaltensmuster präzise und erkennen Abweichungen, um Insider-Bedrohungen frühzeitig zu identifizieren und zu neutralisieren; **Erweiterte Bedrohungsanalyse** verfeinern Bedrohungsmodelle, indem sie Angriffsvektoren simulieren und deren Wahrscheinlichkeit sowie Schwere bewerten, was die Vorhersage und Prävention von Sicherheitsvorfällen verbessert.

Self Organizing Maps (SOM)-Neural Network, auch **Kohonen-Karte** genannt, und wird verwendet im **unüberwachten Lernen** und einem kompetitiven Lernalgorithmus, um komplexe, hochdimensionale Daten durch Clustering und Dimensionsreduktion auf eine niedrigere Dimension zu projizieren, was die Interpretation erleichtert. Eine SOM besteht aus einer Eingabe- und einer Ausgabeschicht. Die Gewichte der Knoten werden zunächst zufällig initialisiert. Für jedes Trainingsbeispiel wird dann der Knoten mit der geringsten Distanz zum Eingabevektor ermittelt, bekannt als **Best Matching Unit (BMU)**. Die SOM passt die Gewichte der BMU und ihrer Nachbarn an, wobei die Anpassungen über die Zeit abnehmen. Dieser Prozess wiederholt sich über mehrere Iterationen, bis die Gewichte optimiert sind und zur Klassifikation neuer Daten verwendet werden können und dieser strukturierte Ansatz ermöglicht es der SOM, Muster in den Daten effizient zu erfassen und zu visualisieren.⁹⁰

vgl.⁸⁹ (Clickworker)
vgl.⁹⁰ (Geeksforgeeks, 2023)

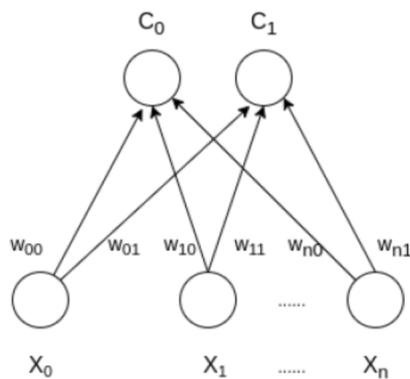


Abbildung 86: Self Organizing Maps

Der Einsatz in SIEM-Anwendungen können sein: **Anomalieerkennung** sind effizient in der Erkennung von Anomalien im Netzwerkverkehr und in Protokolldaten und projizieren hochdimensionale Daten auf eine Karte und machen dadurch ungewöhnliche Muster und potenzielle Bedrohungen sichtbar; **Clustering und Musteranalyse** organisieren Daten in Cluster, ideal für die Analyse großer Datenmengen macht und helfen, Netzwerkaktivitäten und Benutzerverhalten zu klassifizieren und bösartige von normalen Aktivitäten zu unterscheiden; **Visuelle Datenaufbereitung** bieten eine visuelle Darstellung komplexer Sicherheitsdaten, die die Analyse und Interpretation erleichtert sowie Sicherheitsanalysten können so verborgene Muster und Zusammenhänge schnell identifizieren; **Erkennung von Insider-Bedrohungen** und durch die Analyse von Benutzerverhalten über längere Zeiträume können subtile Veränderungen erkennen, die auf Insider-Bedrohungen hindeuten, und so frühzeitig auf interne Sicherheitsrisiken aufmerksam machen; **Optimierte Bedrohungsbewertung** helfen bei der Priorisierung von Bedrohungen, indem die Muster und deren potenzielle Auswirkungen analysieren, was eine gezielte und effektive Reaktion auf die risikoreichsten Bedrohungen ermöglicht.

Deep Belief-Neural Network (DBN) sind fortschrittliche generative Modelle im Deep Learning, die aus mehrschichtigen neuronalen Netzwerken bestehen und diese erkennen und erlernen komplexe Muster in großen Datenmengen, indem sie Informationen Schicht für Schicht verarbeiten. Diese Netzwerke sind besonders nützlich für Anwendungen wie Bild- und Spracherkennung, bei denen hochdimensionale Daten verarbeitet werden müssen. DBNs bestehen aus mehreren Schichten stochastischer Einheiten, oft realisiert durch **Restricted Boltzmann Machines (RBM)**. Jede Schicht extrahiert Merkmale aus den Daten und baut auf den Erkenntnissen der vorherigen Schicht auf, was zu einer tiefen und komplexen Datendarstellung führt. DBNs eignen sich sowohl für **unüberwachtes Lernen**, bei dem sie Muster ohne vorherige Kennzeichnung entdecken, als auch für **überwachtes Lernen**, bei dem sie auf gekennzeichneten Daten trainiert werden. Der Lernprozess in DBNs umfasst zwei Hauptphasen: **Vortraining**, wo jede Schicht unabhängig als RBM trainiert wird, um die Struktur der Daten zu erfassen und diese schichtweise Training hilft, robuste Datendarstellungen zu entwickeln sowie **Feinabstimmung**, diese werden nach dem Vortraining, der Netzwerkparameter, durch Backpropagation für spezifische Aufgaben wie Klassifikation optimiert. Wichtige Konzepte sind **RBM** als Zweischichtige Netzwerke, die die Wahrscheinlichkeitsverteilung der Daten erlernen; **Stochastische Einheiten** erlauben probabilistische Entscheidungen zur Erfassung komplexer Muster; **Greedy-Algorithmus** dienen zum unabhängigen Training und steigert die Effizienz; **Backpropagation** optimiert die Netzwerkleistung für überwachende Aufgaben. DBNs haben durch ihre tiefe Architektur und effizienten Lernmethoden ragen entscheidend bei zur Weiterentwicklung des Deep Learning, insbesondere bei komplexen Aufgaben wie der Bild- und Spracherkennung.⁹¹

vgl.⁹¹ (Geeksforgeeks, 2023)

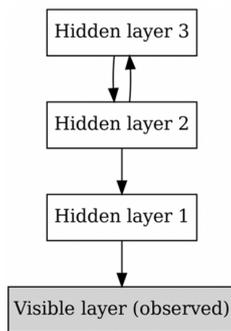


Abbildung 87: Deep Belief Networks

Der Einsatz in SIEM-Anwendungen können sein: **Anomaliedetektion** identifizieren subtile, schwer erkennbare Anomalien im Netzwerkverkehr und in Protokolldaten, indem tiefgehende Muster erlernen, die auf potenzielle Sicherheitsbedrohungen hinweisen; **Bedrohungsklassifikation** ermöglichen eine präzise Klassifikation von Bedrohungen durch die Extraktion komplexer Merkmale aus umfangreichen Datensätzen, was die Effektivität von SIEM-Systemen deutlich verbessert; **Modellierung von Verhalten** modellieren das normale Verhalten von Benutzern und Systemen und erkennen Abweichungen, die auf Insider-Bedrohungen oder unautorisierte Aktivitäten hinweisen könnten; **Datenintegration** integrieren Daten aus verschiedenen Quellen, wie Netzwerkprotokolle und Benutzerverhalten, und bieten so eine umfassende Bedrohungsanalyse in SIEM-Systemen; **Vorhersage von Sicherheitsvorfällen** werden durch das Training auf historischen Daten können zukünftige Sicherheitsvorfälle vorhersagen und ermöglichen so eine proaktive Sicherheitsstrategie.

Restricted Boltzmann Machines-Neural Networks (RBM) wird als **unüberwachtes Lernen** eingesetzt und als generative wird eine Wahrscheinlichkeitsverteilung über Eingabedaten erfasst. RBM bestehen aus zwei Schichten: einer sichtbaren (Eingabeschicht) und einer verborgenen Schicht. Verbindungen zwischen Neuronen derselben Schicht sind nicht erlaubt, was die Reduzierung der Dimensionalität ermöglicht. RBM werden durch kontrastive Divergenz trainiert, einem Verfahren zur Anpassung der Verbindungsgewichte, um die Wahrscheinlichkeit der Trainingsdaten zu maximieren. Nach dem Training können RBM neue Muster generieren und in Bereichen wie Bildverarbeitung, natürliche Sprachverarbeitung und Spracherkennung eingesetzt werden. Hierbei funktionieren neuronale Netzwerke, bei denen alle Neuronen miteinander verbunden sind, jedoch ohne Ausgabeschicht. Sie sind stochastische Modelle, die interne Repräsentationen erlernen und komplexe Probleme lösen können. In einem RBM aktiviert die Eingabeschicht die verborgene Schicht, deren Aktivierungsmuster wiederum die Eingabeschicht rekonstruiert und durch diese Rückkopplungsschleife werden die Verbindungsgewichte angepasst. Dies ermöglicht RBMs, komplexe Muster zu erkennen, wie z. B. in Empfehlungssystemen, wo Nutzerpräferenzen basierend auf früherem Verhalten vorhergesagt werden. RBMs lassen sich in zwei Haupttypen einteilen: **binäre RBM** für binäre Daten und **gauss'sche RBM** für kontinuierliche Daten. Erweiterungen wie **Deep Belief Network** (mehrere RBM-Schichten), **Convolutional RBM** (für Bilder) und **Temporal RBM** (für zeitliche Daten) erlauben den Einsatz in hochdimensionalen und komplexen Datenumgebungen.⁹²

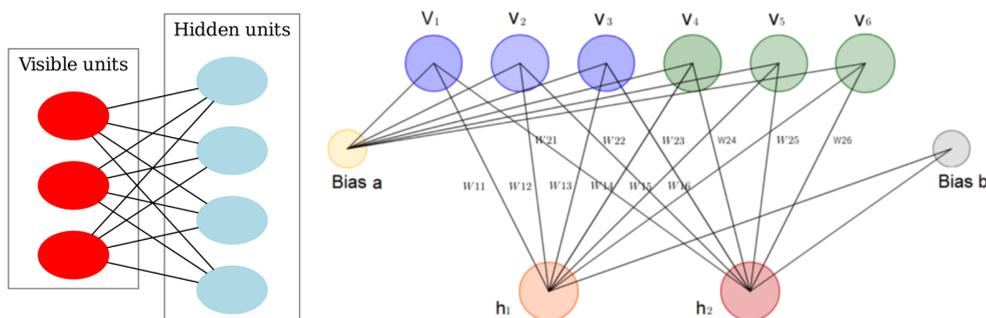


Abbildung 88: Restricted Boltzmann Machines

vgl.⁹² (Geeksforgeeks, 2023)

Der Einsatz in SIEM-Anwendungen können sein: **Präzise Anomaliedetektion** erkennen effektiv Anomalien im Netzwerkverkehr und in Protokolldaten, indem typische Muster erlernen und Abweichungen identifizieren, die auf potenzielle Sicherheitsbedrohungen hinweisen; **Effiziente Merkmalsextraktion** extrahieren wichtige Merkmale aus komplexen Datensätzen und verbessern so die Analyse und Klassifikation von Sicherheitsvorfällen in SIEM-Systemen; **Datenreduktion für effizientere Analyse** reduzieren die Dimension großer Datensätze, indem kompakte Repräsentationen der Daten erstellen, was die Verarbeitung und Analyse in SIEM-Systemen effizienter macht; **Vorbereitung für tiefe Netzwerke** fungieren als Vorstufe für tiefere neuronale Netzwerke wie DBNs, indem Merkmale vorverarbeiten und die Leistung komplexer SIEM-Anwendungen verbessern; **Frühzeitige Erkennung von Insider-Bedrohungen** modellieren normales Benutzerverhalten und erkennen Abweichungen frühzeitig, um Insider-Bedrohungen zu identifizieren, bevor Schaden anrichten können.

Autoencoders-Neural Networks die im **unüberwachten Lernen** eingesetzt werden, um Daten zu komprimieren und wiederherzustellen und diese bestehen aus einem Encoder, der die Eingabedaten in eine komprimierte latente Darstellung überführt, und einem Decoder, der diese Darstellung zur Rekonstruktion der ursprünglichen Daten verwendet. Autoencoders sollen effiziente Darstellungen erlernt werden, die die Eingabedaten möglichst genau rekonstruieren. Ein typischer Autoencoder besteht aus drei Hauptkomponenten: **Encoder**, dieser reduziert die Dimensionalität der Eingabedaten und extrahiert wesentliche Merkmale; **Bottleneck-Schicht**, diese Schicht speichert die komprimierte, latente Darstellung der Daten; **Decoder**, dieser rekonstruiert die ursprünglichen Daten aus der komprimierten Darstellung. Diese Architektur kann je nach Anwendung variieren, z.B. durch den Einsatz von Faltungsschichten für Bilddaten oder rekurrenten Schichten für sequenzielle Daten. Es gibt verschiedene Autoencoder-Varianten, darunter: **Vanilla Autoencoder** mit einfachen Aufbau für Dimensionalitätsreduktion und Merkmalsextraktion; **Variational Autoencoder (VAE)** erzeugt neue Datenproben, die den Trainingsdaten ähneln; **Convolutional Autoencoder** wird speziell für Bildverarbeitung verwendet und nutzt Faltungsschichten; **Rekurrente Autoencoder** verarbeitet sequenzielle Daten mit rekurrenten Schichten; **Denoising Autoencoder** entfernt Rauschen aus den Eingabedaten; **Adversarial Autoencoder (AAE)** kombiniert Autoencoder mit generativen adversarialen Netzwerken (GANs). Autoencoder werden vielseitig eingesetzt, unter anderem für: **Dimensionalitätsreduktion** mit Komprimierung von Daten unter Beibehaltung wesentlicher Merkmale; **Anomalieerkennung** zur Identifikation von Abweichungen in Daten; **Bilderzeugung** zur Erstellung realistischer Bilder aus latenten Darstellungen; **Bildentrauschung** zur Entfernung von Rauschen aus Bildern; **Transfer Learning** zur Nutzung vortrainierter Modelle zur Leistungssteigerung in anderen Aufgaben. Autoencoder stehen vor Herausforderungen wie Überanpassung, Rechenintensität und eingeschränkter Generalisierungsfähigkeit. Dabei ist die Wahl der richtigen Architektur und Regularisierungstechniken entscheidend, um ihre Effektivität in verschiedenen Anwendungen zu maximieren. Trotz dieser Herausforderungen bleiben Autoencoder ein zentraler Bestandteil moderner maschineller Lerntechnologien, mit wachsender Relevanz in vielen Bereichen.⁹³

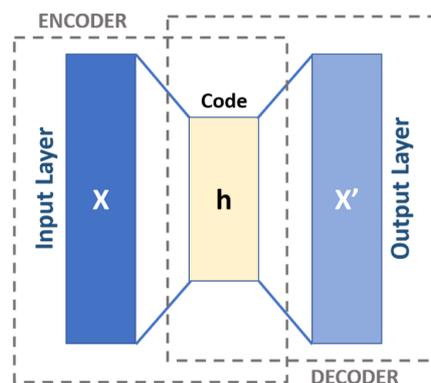


Abbildung 89: Autoencoders

vgl.⁹³ (CAMP, 2023)

Der Einsatz in SIEM-Anwendungen können sein: **Anomaliedetektion** erkennen Anomalien, indem typische Datenmuster rekonstruieren und Abweichungen durch erhöhte Rekonstruktionsfehler aufzeigen und macht diese ideal für die Identifizierung unregelmäßiger Aktivitäten im Netzwerkverkehr oder in Protokolldaten; **Datenkompression und -reduktion** reduzieren die Dimension großer Datenmengen, komprimierte, latente Repräsentationen erzeugen und dadurch wird eine effizientere Speicherung und Verarbeitung in SIEM-Systemen ermöglicht, ohne wesentliche Informationen zu verlieren; **Extraktion relevanter Merkmale** extrahieren bedeutende Merkmale aus komplexen Datensätzen und verbessern so nachfolgende Analyseprozesse wie Bedrohungsclassifikation oder Verhaltensanalyse in SIEM-Systemen; **Vorverarbeitung für komplexere Modelle** dienen als Vorverarbeitungswerkzeug für andere maschinelle Lernmodelle, indem relevante Merkmale extrahieren und Rauschen in den Daten reduzieren, was komplexere Modelle bei der genauen Analyse unterstützt; **Rekonstruktion beschädigter Daten** werden zur Wiederherstellung beschädigter oder unvollständiger Daten eingesetzt und sichern so eine vollständige Analyse in SIEM-Systemen, auch wenn Teile der Daten fehlen oder beschädigt sind.

Im **maschinellen Lernen** werden selbstadaptive Algorithmen verwendet, die sich kontinuierlich verbessern können, ohne dass zusätzliche Eingriffe oder Konfigurationen durch Programmierer erforderlich sind. Dies setzt jedoch eine große Menge qualitativ hochwertiger, strukturierter Daten voraus – oft als Big Data bezeichnet –, die als Trainingsmaterial dienen und dem System ermöglichen, eigenständig zu lernen. Der Lernprozess basiert auf der Zuführung von Trainingsdaten, während die Entscheidungsgrundlagen des Modells durch neue Inputdaten geprüft und verfeinert werden. Feedbackdaten tragen dazu bei, die Leistung weiter zu optimieren, indem sie auf bereits gesammelten Erfahrungen aufbauen. Je mehr Trainingsdaten vorhanden sind, desto leistungsfähiger wird das System, wie z.B. für den Einsatz ML ist die Integration in SIEM-Systeme (Security Information and Event Management), die Muster in Echtzeit zu erkennen, die Anomalien oder verdächtige Aktivitäten erkennen, die eventuelle Sicherheitsverletzungen oder Bedrohungen anzeigen. Dabei unterstützen Maschinelle Lernalgorithmen bei Vorhersagen von Werten in den integrierten Daten, bestimmte Ereignisse werden durch Wahrscheinlichkeitsberechnung ermittelt, Datenclustergroupierungen, Analyse von Sequenzen, Reduktion von signifikanten, Echtzeiterkennungsmustern können potenzielle Sicherheitsvorfälle schneller identifiziert und schnellstmögliche Gegenmaßnahmen von IT-Team eingeleitet werden. Dabei sind kontinuierlichen Anpassung an neue Bedrohungen und Verhaltensweisen verbessert werden in Modelle, um effizienter die Erkennung und Abwehr von Sicherheitsrisiken im IT-Bereich zu behandeln.⁹⁴

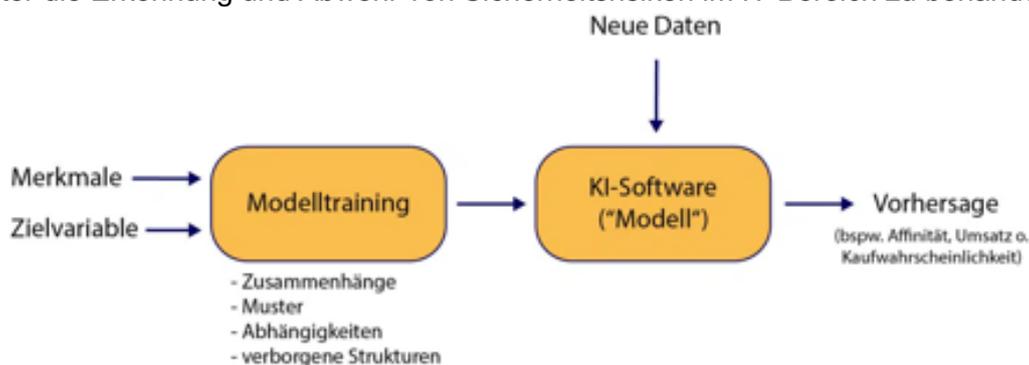


Abbildung 90: Machine Learning nutzt Daten, um Muster und Zusammenhänge in Daten zu identifizieren

3.1.2.1.4. Arten von Algorithmen

Lineare Regression-Algorithmus ist ein Verfahren des überwachten Lernens, das lineare Zusammenhänge in der Funktion zwischen den beobachteten Daten und einem abhängigen Zielwert liegt. Dabei findet die lineare Regression in einer Gerade statt, die die Daten möglichst gut beschreibt, indem diese quadratischen Abstände zwischen den beobachteten Datenpunkten und der Gerade minimiert.

vgl.⁹⁴ (IBM, 2023)

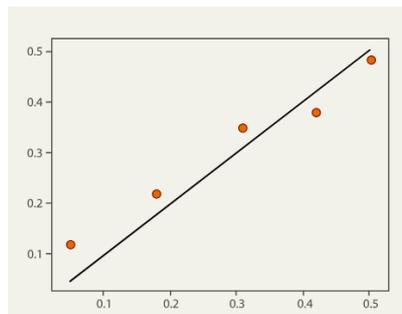


Abbildung 91: Lineare Regression-Algorithmus

$$E_{LSQ}(a, b) = \sum_{x \text{ in Trainingsmenge}} \left((a, b) \cdot \begin{pmatrix} x \\ 1 \end{pmatrix} - F(x) \right)^2$$

E_{LSQ} - kleinsten Quadrate (engl. 'Least Squares', LSQ) minimiert eine Fehlerfunktion E_{LSQ}
a, b - Parameter
 $\sum x$ - in Trainingsdaten

Zur Erreichung dieses Ziels wird das **Least Squares** - Verfahren der kleinsten Quadrate (LSQ) angewendet, das die Fehlerfunktion minimiert und den quadratischen Abstand zwischen den beobachteten Datenpunkten und den Vorhersagen der Geraden beschreibt. Die Methode der kleinsten Quadrate liefert die Koeffizienten für die Geradengleichung, mit denen Vorhersagen für neue Datenpunkte möglich sind. Die praktische Umsetzung der linearen Regression kann durch die Verwendung von Python und dem Paket Numpy erfolgen. Dabei werden die Trainingsdaten geladen, die Fehlerfunktion in Matrix-Vektor-Schreibweise definiert und die Methode der kleinsten Quadrate aufgerufen, um die Koeffizienten der Geraden zu berechnen und so werden die gefundenen Koeffizienten dann verwendet, um Vorhersagen für neue Datenpunkte zu machen. Die lineare Regression ist ein schnelles und zuverlässiges Verfahren, das sowohl auf großen als auch kleinen Datenmengen angewendet werden kann und in verschiedenen Anwendungsbereichen, wie z.B. der Vorhersage von Temperatur oder Preisen, eingesetzt wird.⁹⁵

Logistische Regression-Algorithmus wird im maschinellen Lernen als Klassifizierungsaufgaben verwendet. Ihre Aufgabe besteht darin, die Wahrscheinlichkeit zu schätzen, dass eine Beobachtung zu einer bestimmten Klasse gehört und dieser Algorithmus wird sowohl in der Medizin als auch in der Finanzanalyse sowie im Marketing eingesetzt. Der logistische Regressionsprozess beginnt mit der Anpassung einer Logistischen Kurve an die Daten. Die **Sigmoid-Funktion** transformiert die lineare Kombination der Eingabemerkmale. Die logistische Regression wird durch die Maximierung der **Likelihood-Funktion** bestimmt, wobei verschiedene Optimierungsalgorithmen wie der Gradientenabstieg-Algorithmus verwendet werden können. Die logistische Regression ist ein flexibler und interpretierbarer Algorithmus, der sich hervorragend für binäre Klassifizierungsaufgaben eignet und diese kann auch mit Ansätzen wie "**One-vs-All**" oder "**One-vs-One**" kombiniert werden. Die Qualität des Modells wird oft anhand verschiedener Metriken wie Genauigkeit, Präzision und F1-Score bewertet.⁹⁶

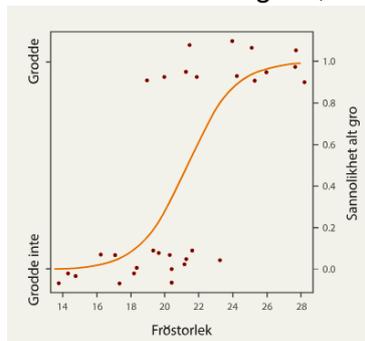


Abbildung 92: Logistische Regression-Algorithmus

vgl.⁹⁵ (Welke)
 vgl.⁹⁶ (AWS)

	Lineare Regression	Logistische Regression
Wie lautet es?	Eine statistische Methode zur Vorhersage eines Ausgabewerts aus einer Reihe von Eingabewerten.	Eine statistische Methode zur Vorhersage der Wahrscheinlichkeit, dass ein Ausgabewert einer bestimmten Kategorie angehört, aus einer Reihe von kategorialen Variablen.
Beziehung	Lineare Beziehung, dargestellt durch eine gerade Linie.	Logistische Beziehung oder sigmoidale Beziehung, dargestellt durch eine S-förmige Kurve.
Gleichung	Linear.	Logarithmisch.
Art des überwachten Lernens	Regression.	Klassifizierung.
Verteilungsart	Normal/Gauß.	Binomisch.
Am besten geeignet für	Aufgaben, die eine vorhergesagte kontinuierliche abhängige Variable aus einer Skala erfordern.	Aufgaben, die eine Vorhersage der Wahrscheinlichkeit des Auftretens einer kategorialen abhängigen Variable aus einem festen Satz von Kategorien erfordern.

Abbildung 93: Vergleich lineare Regression vs. logistische Regression

Naïve Bayes-Naive Bayes-Klassifikatoren-Algorithmus ist ein einfacher, aber leistungsstarker Algorithmus, der auf dem **Bayes-Theorem** basiert und zur Klassifikation verwendet wird und berechnet die Wahrscheinlichkeit, dass ein Datenpunkt zu einer bestimmten Klasse gehört, unter der Annahme, dass die Merkmale unabhängig voneinander sind. Trotz dieser Vereinfachung liefern naive Bayes-Klassifikatoren oft beeindruckende Ergebnisse, insbesondere bei Textklassifikation und Spam-Erkennung. Varianten wie **Gaussian**, **Multinomial** und **Bernoulli Naive Bayes** sind für unterschiedliche Datentypen optimiert. Die Vorteile des Modells liegen in seiner Einfachheit, Geschwindigkeit und Effizienz, besonders bei großen Merkmalsräumen und kann allerdings die Annahme der Unabhängigkeit von Merkmalen zu Einschränkungen führen, wenn Merkmale stark korrelieren.⁹⁷

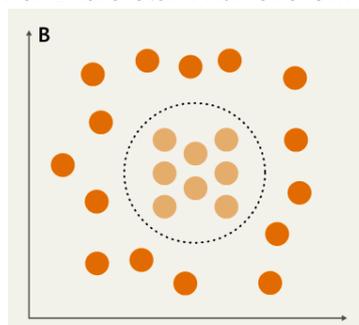


Abbildung 94: Naïve Bayes-Naive Bayes-Klassifikatoren-Algorithmus

Support Vector Machine-Algorithmus (SVM) ist eine komplexe Klassifizierungs-, Regressions- und Ausreißererkenngungsprobleme löst. Dieser ermittelt Hyperplane, die Datenpunkte verschiedener Klassen optimal trennen und nutzt Kernelmethoden, um Daten in höhere Dimensionen zu transformieren und diese leichter zu trennen. Dies funktioniert, indem man Trennung zwischen verschiedenen Kategorien von Datenpunkten in einem mehrdimensionalen Raum findet. Diese Trennung wird durch eine sogenannte Hyperplane realisiert, die so positioniert wird, dass der Abstand zu den nächstgelegenen Datenpunkten, den sogenannten Supportvektoren, maximiert wird. Sie werden in vielen Bereichen eingesetzt, einschließlich der Geo-Sound-Verfolgung, seismischen Flüssigkeitspotenzialbewertung, Proteinerkennung, Datenklassifizierung, Gesichtserkennung, Oberflächentexturklassifizierung, Handschriftenerkennung, Spracherkennung, Steganographieerkennung und Krebsdiagnose.⁹⁸

vgl.⁹⁷ (IBM)
vgl.⁹⁸ (Kanade, 2022)

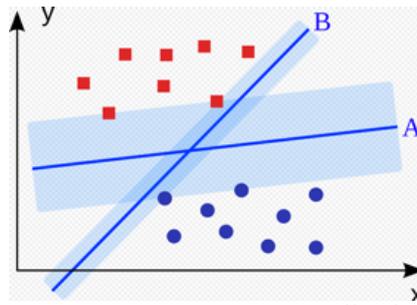


Abbildung 95: - Support Vector Machine-Algorithmus (SVM) Algorithmus

Entscheidungsstruktur-Algorithmen, auch als Entscheidungsbäume bezeichnet, werden im ML verwendet zur Datenanalyse und zeichnet Entscheidungsprozesse auf, indem die Daten schrittweise in immer kleinere, mindestens zwei gleichartige Gruppen aufteilen. Dabei werden die Segmentierung in Knotenpunkten des Baums aufgeteilt und auf spezifischen Entscheidungsmerkmalen aufgespalten werden und die **Wenn-Dann-Regeln** Daten entsprechend von den wichtigsten Unterscheidungsmerkmalen zwischen den Datenpunkten zu differenzieren.⁹⁹

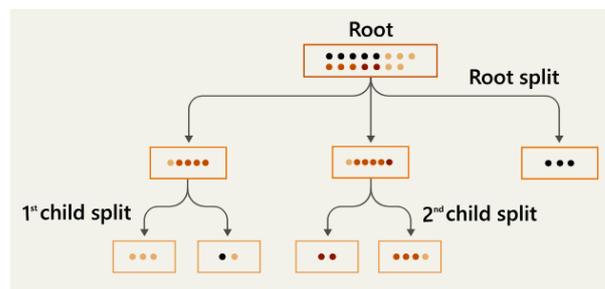


Abbildung 96: Entscheidungsstruktur-Algorithmen

K-Nächste-Nachbarn-Algorithmus (KNN oder k-NN) ist ein nichtparametrischer, überwachter Lernklassifikator mit dem Konzept das auf der Nähe basiert, macht Vorhersagen über Klassenzugehörigkeit, kontinuierlichen Werte eines einzelnen Datenpunktes. Die Ähnlichkeit der Datenpunkte werden bestimmt, zu welchen Klassen ein neuen Datenpunkt zugeordnet werden, um das Klassifikationsproblem zu lösen, erfolgt die Klassenzuweisung durch eine Mehrheitsabstimmung von den K-Nächsten-Nachbarn des betrachteten Datenpunktes und dann eine Klasse zugewiesen wird, die einen bestimmten Prozentsatz der Stimmen übersteigt, die mehr als zwei Klassen besitzen. Der Durchschnitt der K-Nächsten-Nachbarn verwendet das Regressionsproblem, um kontinuierliche Vorhersagen von Datenpunkten zu treffen und eine numerische Schätzung zu bewerten.

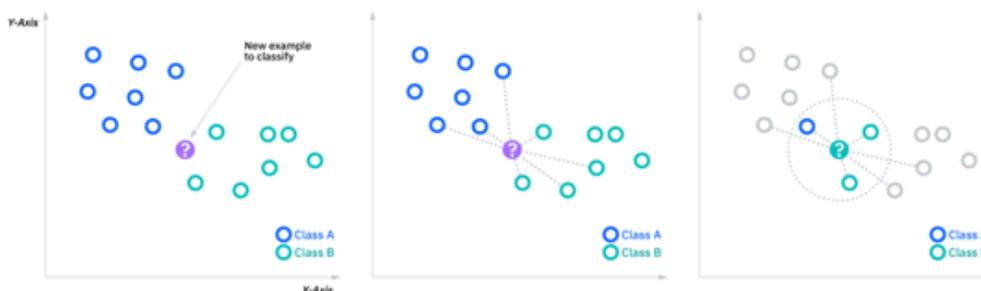


Abbildung 97: KNN-Diagramm

Der KNN-Algorithmus gehört zur Familie der „**Lazy Learning**“-Modelle, d.h. dass dieser keinen separaten Trainingsprozess durchläuft und stattdessen alle Trainingsdaten im Speicher behält und dies macht den Algorithmus einfach, aber auch speicherintensiv und ineffizient für große Datensätze, da seine Leistung mit zunehmender Datengröße abnimmt. Der KNN-Algorithmus wird meist nur noch für einfache Empfehlungssysteme, Mustererkennung, Data Mining und andere Anwendungen verwendet.¹⁰⁰

vgl.⁹⁹ (IBM, 2024)
vgl.¹⁰⁰ (IBM)

Clustering-Algorithmus bezieht sich "**Clustering**" auf eine Methode des maschinellen Lernens, bei der ähnliche Datenpunkte in Gruppen oder "Cluster" zusammengefasst werden. Das Clustering dient dazu, intrinsische (innerliche) Strukturen in einem Datensatz zu entdecken, indem Datenpunkte mit ähnlichen Merkmalen oder Verhaltensweisen ermittelt und gruppiert werden. Das Clustering ist eine Form des **unüberwachten Lernens**, da es keine vorherigen Labels oder Kategorien für die Datenpunkte gibt. Stattdessen versucht der Clustering-Algorithmus, natürliche Gruppierungen oder Muster in den Daten finden, ohne auf externe Informationen angewiesen zu sein. Die Verwendung von Clustering in der KI sind vielfältig, wie **Mustererkennung** mit Clustering-Algorithmen können dabei helfen, verborgene Muster oder Strukturen in großen Datensätzen zu identifizieren, die menschlichen Analysten möglicherweise nicht erkennen; **Segmentierung** können durch Clustering Daten in sinnvolle Segmente oder Gruppen unterteilt werden, um Kundenbasis zu verstehen, Produkte zu personalisieren oder Marketingstrategien zu entwickeln; **Anomalieerkennung** mit Clustering kann auch zur Erkennung von Anomalien oder Ausreißern in Daten verwendet werden, indem es Punkte ermittelt werden, die nicht gut zu den vorhandenen Clustern passen sowie **Vorverarbeitung** in einigen Fällen dient Clustering als Vorverarbeitungsschritt für andere maschinelle Lernverfahren, indem es die Dimensionalität reduziert oder die Daten in Form von Clustern darstellt, um die Modellleistung zu verbessern.

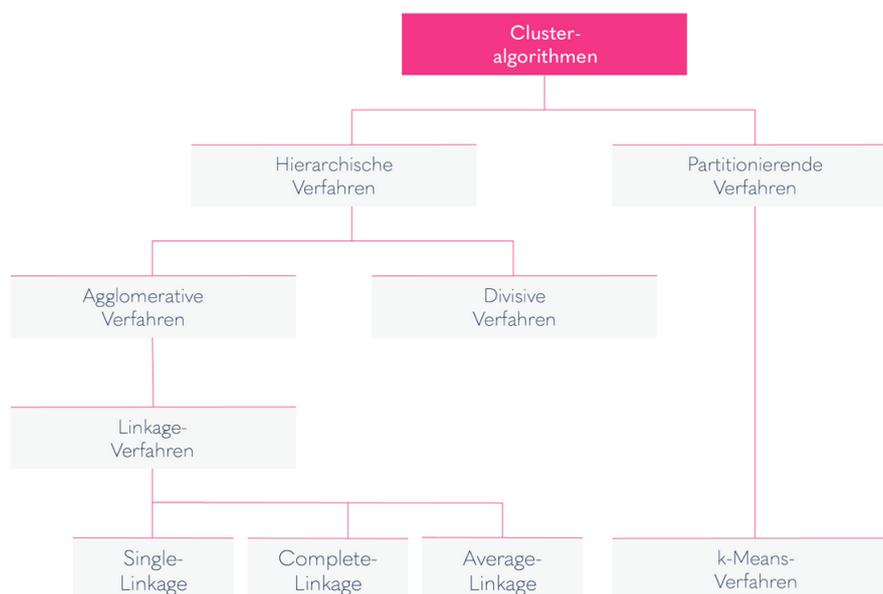


Abbildung 98: Clustering-Algorithmus

Clusteranalysen lassen sich in hierarchische und partitionierende Verfahren unterteilen. **Hierarchische Methoden**, wie agglomerative Verfahren, beginnen mit kleinen Clustern, die schrittweise zu größeren zusammengeführt werden. **Partitionierende Verfahren** sind wie k-Means-Algorithmus und optimieren die Struktur bereits klassifizierter Datensätze, indem sie die Daten in einer definierten Anzahl von Clustern neu anordnen. Zur Clusterbildung können verschiedene **Linkage-Methoden** genutzt werden:

- **Single-Linkage:** Minimale Distanz zwischen Clustern.
- **Complete-Linkage:** Maximale Distanz zwischen den entferntesten Punkten.
- **Average-Linkage:** Durchschnittliche Distanz zwischen allen Punkten.

K-Means Clustering ist eine Methode zur Gruppierung von Datenpunkten in Untergruppen, wobei ähnliche Datenpunkte in derselben Gruppe landen sollen. Der k-Means Algorithmus ist ein häufig verwendetes Clusteringverfahren und zunächst werden die k-zufällige Clusterzentren ausgewählt. Anschließend werden die Datenpunkte den am nächsten liegenden Clusterzentren zugeordnet und danach werden die Clusterzentren neu berechnet, indem der Mittelpunkt aller Punkte in einem Cluster bestimmt wird. Dieser Prozess wird solange wiederholt, bis sich die Clusterzentren nicht mehr oder nur minimal verändern.

Um den k-Means Algorithmus zu verstehen sind fünf Schritte zu durchlaufen:

1. **Bestimmung der Anzahl der Cluster (k)** - Entscheidung, wie viele Cluster in einen Datensatz gelegt werden.
2. **Initialisierung der Clusterzentren** - Auswahl treffen der zufällig k-Datenpunkte, die als Anfangspositionen im Clusterzentren, auch Zentroide genannt werden und die Wahl dieser Anfangszentroide wird die Leistung des Algorithmus beeinflussen, daher ist es wichtig, verschiedene Initialisierungsmethoden zu testen.
3. **Zuordnung der Datenpunkte zu den Clustern** – Zuordnung der Datenpunkte zu dem Cluster, dessen Zentroid ihm am nächsten liegt.
4. **Neuberechnung der Clusterzentren** – Berechnung zu jeden Cluster wird ein neues Zentrum ermittelt, das dem Punkt im Cluster entspricht, der den geringsten Gesamtabstand zu allen anderen Punkten im Cluster hat.
5. **Überprüfung auf Konvergenz** – Überprüfung vom Algorithmus, ob sich die Position der Clusterzentren seit der letzten Wiederholung wesentlich verändert hat, wenn aber nicht, wird der Algorithmus beendet und geht andernfalls zurück zum Schritt 3 und ordnet die Datenpunkte erneut den Clustern zu.

Diese Schritte werden solange wiederholt, bis die Clusterzentren stabil sind und der Algorithmus übereinstimmend ist.

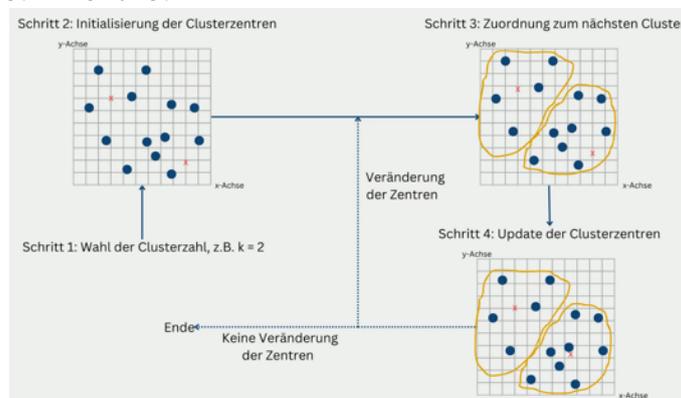


Abbildung 99: k-Means Clustering Prozess

Der k-Means Algorithmus findet in verschiedenen Anwendungen, darunter Bildsegmentierung, Kundensegmentierung, Anomalieerkennung, genomisches Clustering und Analyse sozialer Netzwerke. Die richtige Anzahl von Clustern zu wählen, kann eine Herausforderung sein, aber die Ellbogen-Methode kann dabei helfen, den optimalen k-Wert zu bestimmen. Bei der Bewertung der Leistung eines k-Means Clusters können Kennzahlen, wie z.B. die Summe der Quadrate innerhalb eines Clusters, der Silhouetten-Wert, der Davies-Bouldin-Index und der Calinski-Harabasz-Index verwendet werden. Trotz seiner Nützlichkeit hat der k-Means Algorithmus auch Grenzen, wie z.B. die Notwendigkeit, die Anzahl der Cluster im Voraus festzulegen, die Empfindlichkeit gegenüber der Initialisierung der Clusterzentren, die Annahme von kugelförmigen Clustern und die Anfälligkeit gegenüber Ausreißern.¹⁰¹

Density-Based Spatial Clustering of Applications with Noise (DBSCAN-Algorithmus) ist eine leistungsfähige Methode zur Clusteranalyse von Datenpunkten, die auf ihrer Dichte in einem Raum basiert. Im Gegensatz zu k-Means ist DBSCAN ein datengetriebener Algorithmus, der automatisch die Anzahl der Cluster bestimmt und auch Ausreißer erkennen kann. Seine Funktionsweise bezieht sich auf die Datenpunkte, die in Kernpunkte, Grenzpunkte und Ausreißer eingeteilt werden. **Kernpunkte** besitzen eine bestimmte Mindestanzahl von Nachbarn innerhalb eines bestimmten Radius. **Grenzpunkte** haben weniger Nachbarn und sind von einem Kernpunkt erreichbar. **Ausreißer** sind weder Kernpunkte noch Grenzpunkte, da der Benutzer Parameter festlegen muss: die **Epsilon (ϵ)**, der **Radius um jeden Datenpunkt**, der innerhalb dessen Nachbarn gesucht werden, sowie die **MinPts**, die **Mindestanzahl von Nachbarn**, die einen Punkt haben muss, um als Kernpunkt betrachtet zu werden. Die Clusterbildung beginnt mit einem beliebigen, nicht besuchten Punkt und der

vgl.¹⁰¹ (CAMP, 2022)

Algorithmus findet alle Punkte in der ϵ -Nachbarschaft dieses Punktes. Wenn der Punkt selbst ein **Kernpunkt** ist, wird ein neues Cluster gestartet, und alle erreichbaren Punkte werden diesem Cluster zugeordnet. **Grenzpunkte** werden dem Cluster zugeordnet, zu dem der Kernpunkt gehört. **Ausreißer** werden in diesen Prozess solange wiederholt, bis diese zu einer zugehörigen Gruppe gruppiert werden.

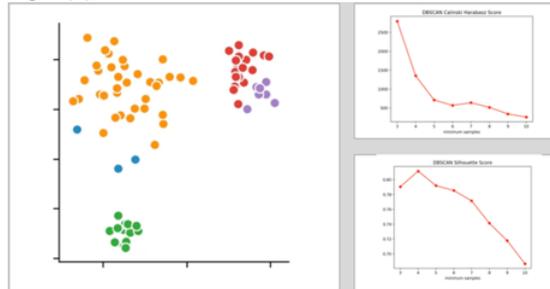


Abbildung 100: Ergebnisse unserer Clusteranalyse-Beispiels. Clusterbildung mit dem DBSCAN-Algorithmus. Auswertung der gefundenen Cluster mit dem Calinski-Harabasz-Index und der Silhouttenmethode

DBSCAN bietet einige Vorteile gegenüber k-Means, wie z.B. die automatische Bestimmung der Anzahl von Clustern und die Fähigkeit, Cluster beliebiger Form zu identifizieren. Dabei werden die Parameter ϵ und **MinPts** sorgfältig ausgewählt und die Skalierung der Daten zu berücksichtigt. DBSCAN Anwendung erfolgt bei denen die Form und Dichte der Cluster variieren können, wie die Segmentierung von Bildern, die Erkennung von Anomalien in Daten und die Gruppierung von Geodatenpunkten.¹⁰²

Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) ist eine leistungsstarke Methode zum Clustern von Datenpunkten, bei der die kleinste Clustergröße variiert wird und dabei gibt diese kleinste Clustergröße an, wie viele Punkte bei einem hierarchischen Verfahren erforderlich sind, damit isolierte Datenpunkte als neue Cluster betrachtet werden. Durch die Anwendung von Metriken wie dem **Calinski-Harabasz-Index** und der **Silhouttenmethode** können optimale Clustergrößen bestimmt werden.

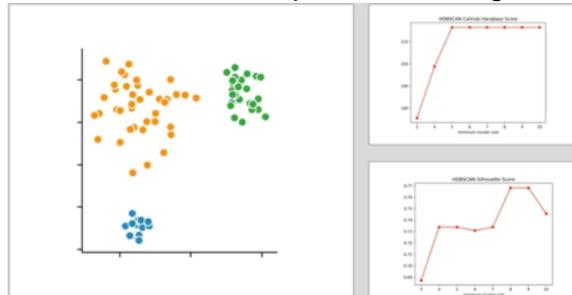


Abbildung 101: Clusterbildung mit dem HDBSCAN-Algorithmus. Auswertung der gefundenen Cluster mit dem Calinski-Harabasz-Index und der Silhouttenmethode

Dabei kann die optimale Clustergröße zwischen vier und fünf Punkten liegen und es wird entscheiden sich für die Vier und erhält dann entsprechend ein Clustering-Result. Dies verdeutlicht, wie wichtig es ist, verschiedene Metriken zur Validierung der Clustering-Ergebnisse heranzuziehen und sie miteinander zu vergleichen. Da es viele verschiedene Clustering-Verfahren gibt, die unterschiedliche Lösungen liefern können, ist es entscheidend, den geeignetsten Algorithmus für den spezifischen Anwendungsfall zu finden und alle weiteren Parameter entsprechend anzupassen, sodass die Überprüfung der Richtigkeit von Clustering-Result oft eine Herausforderung sein kann. Die Validierungsmetriken sind nur bedingt geeignet und können je nach Daten und Anwendungsfall unterschiedliche Ergebnisse liefern und deshalb ist es ratsam, eine Vielzahl von Metriken zu verwenden und die Ergebnisse kritisch zu bewerten, um sicherzustellen, dass das gewählte Clustering-Verfahren die bestmöglichen Ergebnisse für die gegebene Problemstellung liefert.¹⁰³

vgl.¹⁰² (Marzell, 2021)

vgl.¹⁰³ (Marzell, 2021)

Hierarchisches Clustering ist ein leistungsfähiges Verfahren zur Datenanalyse, das die hierarchische Strukturierung von Datenpunkten ermöglicht mit dem Ziel, die Beziehungen zwischen den Daten in Form einer Hierarchie abzubilden und diese in entsprechende Gruppen zu unterteilen. Es existieren zwei grundlegende Ansätze: der **Top-Down-Ansatz** (divisives clustering) und der **Bottom-Up-Ansatz** (agglomeratives clustering). Beim **Top-Down-Ansatz** beginnt der Prozess entweder mit einem einzigen umfassenden Cluster, das alle Datenpunkte enthält, oder mit einzelnen Clustern für jeden Datenpunkt. Das anfängliche große Cluster wird schrittweise in kleinere Einheiten zerlegt, indem jeweils der am weitesten entfernte Punkt als neues Clusterzentrum identifiziert wird und solange wird der Aufteilungsprozess fortgesetzt, bis die gewünschte Anzahl von Clustern erreicht ist. Der **Bottom-Up-Ansatz** verfolgt den gegensätzlichen Weg, es werden zunächst alle Datenpunkte als individuelle Cluster behandelt und diese dann sukzessive basierend auf ihrer räumlichen Nähe zu größeren Clustern zusammengeführt. Dieser Fusionsprozess wird solange fortgeführt, bis entweder die vorgegebene Clusteranzahl erreicht ist oder sämtliche Daten in einem einzigen Cluster vereint sind.¹⁰⁴

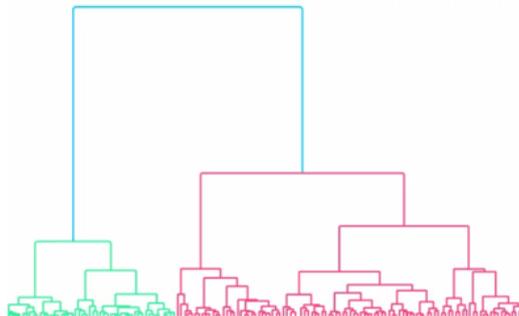


Abbildung 102: beispielhafte Darstellung eines hierarchischen Clusterings beim Machine Learning

Hierarchisches Clustering erzeugt eine **Baumstruktur**, die die hierarchischen Beziehungen zwischen Clustern visualisiert und diese besonders geeignet sind für Daten mit einer natürlichen Hierarchie, wie z.B. in Taxonomien vorkommen. Zu den gängigen Algorithmen des hierarchischen Clustering gehören **BIRCH**, **CURE**, **ROCK** und **Chameleon**. Trotz seiner Vorteile, wie der Fähigkeit, Datensätze mit verschiedenen Merkmalstypen zu verarbeiten und die hierarchischen Beziehungen der Cluster klar darzustellen, weist das hierarchische Clustering auch einige Nachteile auf, wie z.B. die vergleichsweise hohe Zeitkomplexität und die Notwendigkeit, die Anzahl der Cluster im Voraus festzulegen.¹⁰⁵

Weitere Clustering Algorithmen sind **Partitionierendes Clustering** (centroid-based/schwerpunktbasiert), **Dichtebasiertes Clustering** (density-based) oder **Verteilungsbasiertes Clustering** (distribution-based), die unter¹⁰⁶ näher beschrieben sind.

Random-Forest-Algorithmus ist eine leistungsfähige Methode zur Verbesserung von Vorhersagemodellen und basiert auf dem Konzept des **Ensemble-Lernens**, bei dem mehrere Entscheidungsbäume kombiniert werden, um genauere und robustere Vorhersagen zu erzielen. Im Gegensatz zu einem einzelnen Entscheidungsbaum, der anfällig für Überanpassung ist, arbeitet ein Random Forest mit einer Vielzahl von Bäumen, die unabhängig voneinander trainiert werden und dabei wird das Risiko der Überanpassung reduziert, da die Bäume unterschiedliche Muster im Datensatz erkennen können. Der Random-Forest-Algorithmus verwendet zwei Haupttechniken: das **Bagging** und das **Feature Randomness**. Beim **Bagging Randomness** wird zufällige Untergruppen von Daten aus dem Trainingsdatensatz gezogen, um jeden Baum zu trainieren und diese werden dann kombiniert, um eine konsolidierte Vorhersage zu treffen. **Feature Randomness** bezieht sich darauf, dass für jeden Baum nur eine zufällige Teilmenge der Merkmale verwendet wird, um die Vielfalt der Bäume zu erhöhen und die Korrelation zwischen ihnen zu verringern.

vgl.¹⁰⁴ (Marzell, 2021)

vgl.¹⁰⁵ (Marzell, 2021)

vgl.¹⁰⁶ (Marzell, 2021)

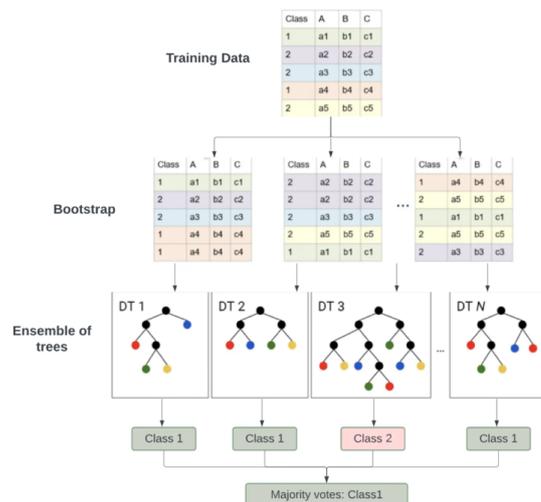


Abbildung 103: Random-Forest-Algorithmus

Der Random-Forest-Algorithmus verbessert die Vorhersagegenauigkeit, hat ein geringeres Überanpassungsrisiko, mehr Flexibilität zur Verwendung von verschiedenen Datentypen, mehr Task für Regressions- und Klassifizierungs und der einfachen Merkmalsbedeutungsbestimmung. In SIEM kann der Random-Forest-Algorithmus besser eine Muster- und Anomalieerkennung in großen Datensätzen analysieren und benutzt die präzise Echtzeitvorhersagen von potenziellen Sicherheitsrisiken und somit eine schnellere Risikenreaktion im Unternehmen.¹⁰⁷

AdaBoost (Adaptive Boosting) und **Gradient Boosting** sind fortschrittliche Ensemble-Methoden im ML, mit den Ziel Vorhersagegenauigkeit durch die Kombination mehrerer schwacher Modelle zu steigern.

AdaBoost trainiert Modelle sequentiell, wobei jedes neue Modell sich auf die Datenpunkte konzentriert, die zuvor falsch klassifiziert wurden. Dabei werden die Gewichte der Datenpunkte so angepasst, dass schwierig zu klassifizierende Fälle stärker berücksichtigt werden und am Ende kombiniert AdaBoost alle Modelle, wobei stärkere Modelle mehr Einfluss auf die Vorhersage haben.

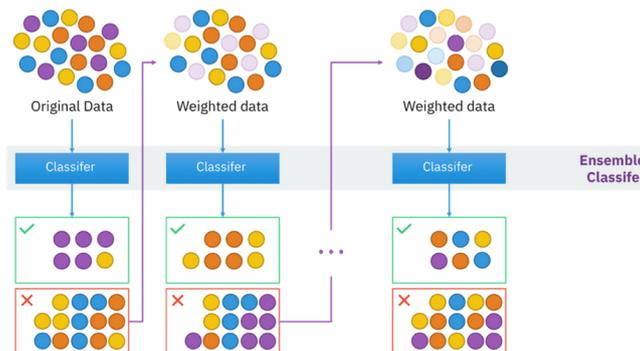


Abbildung 104: AdaBoost

Gradient Boosting verbessert die Modelle ständig bis in jeder Runde der Fehler der vorherigen Modelle durch Gradientenminimierung korrigiert ist. Neue Modelle werden so trainiert, dass diese die verbleibenden Fehler schrittweise reduzieren, was besonders bei komplexen, nichtlinearen Problemen effektiv ist.

AdaBoost erhöht die Gewichtung schwer zu klassifizierender Datenpunkte, während Gradient Boosting Fehler direkt durch Gradientenminimierung optimiert. Dabei sind beide Methoden sehr leistungsstark, wobei AdaBoost für seine Robustheit bekannt ist und Gradient Boosting sich durch seine Fähigkeit auszeichnet, komplexe Zusammenhänge zu modellieren. Die Wahl zwischen den beiden hängt von den spezifischen Anforderungen der Anwendung und den Eigenschaften der Daten ab.¹⁰⁸

vgl.¹⁰⁷ (IBM)
vgl.¹⁰⁸ (Geeksforgeeks, 2023)

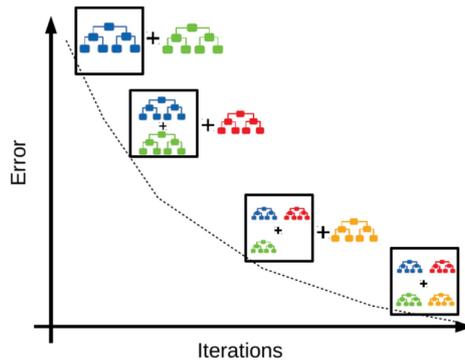


Abbildung 105: Gradient Boosting-Algorithmus

LightGBM ist ein Framework für Gradienten-Boosting, das auf baumbasierten Lernalgorithmen agiert und es zeichnet sich durch hohe Effizienz, schnelle Trainingsgeschwindigkeit und geringen Speicherbedarf aus. Es wurde entwickelt, um die Leistung von Gradient-Boosting-Modellen zu maximieren und den Ressourcenverbrauch zu minimieren und LightGBM ist besonders gut für die Verarbeitung großer Datensätze und das Lernen in verteilten Umgebungen geeignet. Durch die Unterstützung von parallelem und GPU-basiertem Lernen beschleunigt LightGBM den Trainingsprozess erheblich und bietet eine Vielzahl von Konfigurationsoptionen zur Anpassung und Optimierung von Modellen.¹⁰⁹

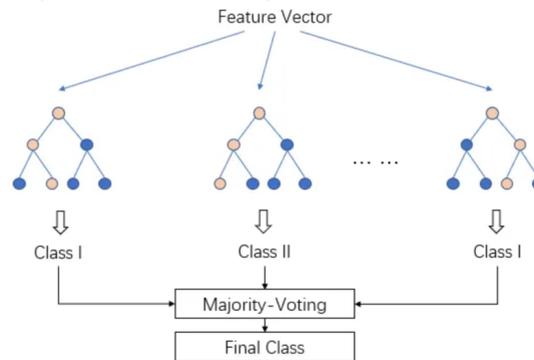


Abbildung 106: LightGBM

Im SIEM-Kontext kann LightGBM verwendet werden, um die Effizienz und Genauigkeit der Ereigniserkennung zu verbessern und Dank seiner schnellen Trainingsgeschwindigkeit und des geringen Speicherbedarfs kann LightGBM verdächtige Aktivitäten oder potenzielle Sicherheitsbedrohungen in großen Datenmengen erkennen und Anomalien in Echtzeit ermitteln. Durch die Entwicklung prädiktiver Modelle auf der Grundlage historischer Sicherheitsdaten ermöglicht es LightGBM, potenzielle Schwachstellen oder Angriffsvektoren frühzeitig zu erkennen und proaktiv darauf zu reagieren.

CatBoost ist ein leistungsstarker Algorithmus für Gradienten-Boosting auf Entscheidungsbäumen, der von Forschern und Ingenieuren bei Yandex entwickelt wurde. Als **Open-Source-Tool** bietet CatBoost eine breite Palette von Anwendungsmöglichkeiten in verschiedenen Branchen und Domänen und zeichnet sich besonders durch seine Fähigkeit aus, mit unstrukturierten Daten umzugehen und komplexe Muster in großen Datensätzen zu erkennen. Anwendungen sind, z.B. Suchmaschinenoptimierung, Empfehlungssysteme, persönliche Assistenten, selbstfahrende Autos und Wettervorhersagen.¹¹⁰

vgl.¹⁰⁹ (LightGBM)
vgl.¹¹⁰ (CatBoost, 2024)

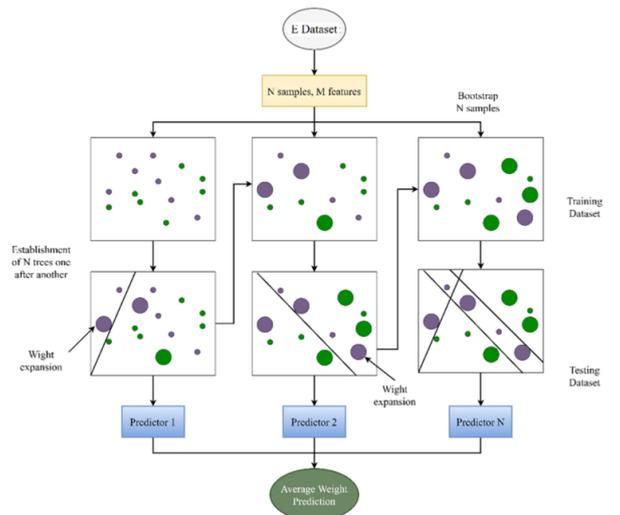


Abbildung 107: CatBoost

Im SIEM-Kontext kann CatBoost dazu beitragen, die Effizienz und Genauigkeit der Ereigniserkennung zu verbessern, indem es verdächtige Aktivitäten oder potenzielle Sicherheitsbedrohungen in Echtzeit ermittelt und Anomalien erkennt bei der Analyse verschiedener Datenquellen wie Netzwerkprotokollen, Logdateien und Endpunktaktivitäten kann CatBoost prädiktive Modelle entwickeln, um zukünftige Sicherheitsvorfälle vorherzusagen und direkt auf potenzielle Bedrohungen zu reagieren.

XGBoost ist eine leistungsstarke Bibliothek für Gradienten-Boosting, die für ihre Effizienz, Flexibilität und Portabilität steht und als optimierte und verteilte Lösung bietet dieser eine Reihe von Funktionen und Vorteilen für die datenwissenschaftliche Gemeinschaft, die durch die ML Implementierung unter dem **Gradienten-Boosting-Framework**, eine schnelle und genaue Lösung komplexer datenwissenschaftlicher Probleme löst.¹¹¹

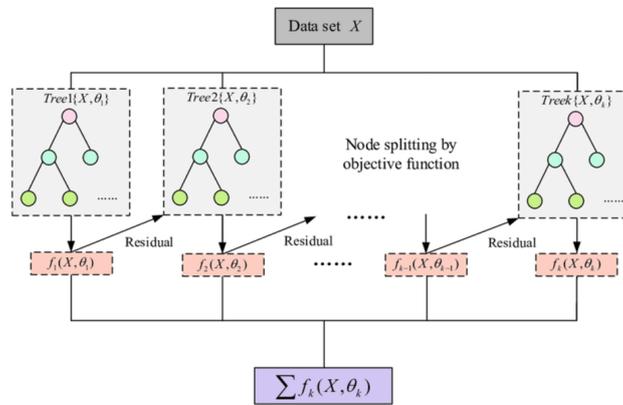


Abbildung 108: XGBoost

Im SIEM-Kontext kann XGBoost verwendet werden, um verdächtige Aktivitäten oder potenzielle Sicherheitsbedrohungen zu identifizieren und Anomalien in großen Datenmengen zu erkennen. Durch die Analyse verschiedener Sicherheitsdatenquellen und die Unterstützung verteilter Umgebungen kann XGBoost die Reaktionszeiten auf Sicherheitsvorfälle verkürzen und die Effizienz der Incident Response zu verbessern. Hierbei werden vorhersagende Modelle von XGBoost genutzt, um historische Sicherheitsdaten, potenzielle Schwachstellen oder Angriffsvektoren frühzeitig zu erkennen und voraussehbar darauf zu reagieren. Bei der Entwicklung und Implementierung von ML sind mehrere Herausforderungen zu beachten:¹¹² **Datenqualität und -verfügbarkeit** sind entscheidend für den Erfolg von ML-Systemen, wobei schlechte Datenquellen zu ungenauen Ergebnissen führen und dies erfordern oft eine aufwendige Vorverarbeitung erfordert.

vgl.¹¹¹ (XGBoost)
vgl.¹¹² (Circle)

Ethik in der ML-Systeme können ethisch bedenkliche Informationen verstärken und zu unfairen Entscheidungen führen und somit sollte sorgfältige, genauere Überwachung und Kontrolle erforderlich sein.

Ressourcenbedarf im Betrieb von ML-Systemen bedarf enorme Rechenleistung mit erheblichen Energieverbrauch und sollte auf Nachhaltigkeit ausgelegt werden.

Overfitting und Underfitting in ML-Modelle können entweder zu viele oder zu wenige Muster in den Daten zu erfassen und sollten dabei die richtige Balance finden, um das volle Potenzial von ML-Systemen auszuschöpfen.

3.1.3. Deep Learning

Deep Learning (DL) ist eine spezielle Form des maschinellen Lernens und bildet eine Teilmenge davon. Im Gegensatz zu herkömmlichen Ansätzen erfordert es keine strukturierten Daten und kann eigenständig lernen. Der Begriff "Deep" im Deep Learning bezieht sich auf die Tiefe der Schichten in einem neuronalen Netz, da ein neuronales Netzwerk mit mehr als drei Schichten – einschließlich Eingabe- und Ausgabeschicht – wird als Deep-Learning-Algorithmus bezeichnet, während Netzwerke mit nur zwei oder drei Schichten als einfache neuronale Netze gelten. KI-Systeme, die auf Deep Learning basieren, können eine breitere Palette an Datenquellen verarbeiten und erfordern dabei weniger Vorverarbeitung durch den Menschen. Dadurch liefern sie häufig genauere Ergebnisse als herkömmliche maschinelle Lernverfahren. Deep Learning nutzt spezielle neuronale Netzwerke, die in der Lage sind, sehr große Mengen an Eingabedaten zu verarbeiten und diese über mehrere Schichten hinweg zu analysieren. Durch die Optimierung dieser Netzwerke erfolgt durch tiefere interne Strukturen, die es ermöglichen, komplexe Muster und Korrelationen zwischen den Datenpunkten zu erkennen und das System lernt aus seinen eigenen Erfahrungen und kann neue Input-Daten mit bereits vorhandenen Informationen in Beziehung setzen. Dabei setzt es zunehmend komplexere Konzepte aus einfacheren Elementen zusammen und entwickelt dadurch ein immer tiefgehendes Verständnis der Daten.¹¹³

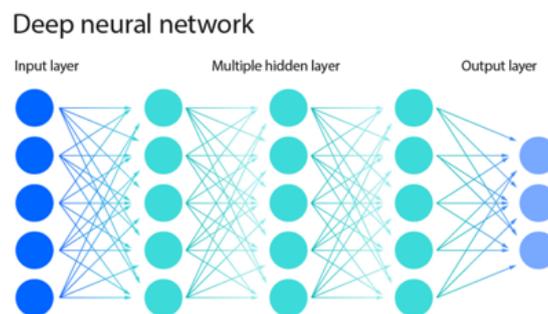


Abbildung 109: Deep neural network

SIEMs Einsatzmöglichkeiten in Deep Learning sind: **Anomalieerkennung** Modelle wie Autoencoder und RNNs lernen normale Verhaltensmuster und identifizieren Abweichungen, die auf Sicherheitsbedrohungen hinweisen könnten, ohne auf vordefinierte Regeln angewiesen zu sein; **Bedrohungserkennung und -korrelation** ermöglicht die automatische Korrelation von Daten aus verschiedenen Quellen und das Erkennen komplexer Angriffsmuster, die von herkömmlichen Methoden oft übersehen werden; **Intrusion Detection** durch den Einsatz von CNNs und LSTMs können SIEM-Systeme sowohl bekannte als auch neuartige Angriffe effektiver erkennen; **Verhaltensanalyse** (UEBA) verbessern die Erkennung von Insider-Bedrohungen durch das Erlernen normaler Benutzerverhalten und die Identifikation subtiler Abweichungen; **Ereignisklassifizierung** durch NLP-Modelle wie Transformer ermöglichen eine genauere Klassifizierung und Priorisierung von Sicherheitsvorfällen durch die Analyse von Protokollen und Warnungen; **Reduktion von Fehlalarmen** lernt DL aus der Rückmeldung der Analysten und hilft, die Anzahl der Fehlalarme zu verringern, indem es sich auf tatsächlich verdächtige Aktivitäten konzentriert; **Phishing-Erkennung** durch DL verbessert die Erkennung von Phishing-Angriffen durch die Analyse von E-Mail-Inhalten und Anhängen, was über die Möglichkeiten traditioneller Methoden hinausgeht; **Automatisierung der Vor-**

vgl.¹¹³ (IBM, 2023)

fallreaktion durch DL und Reinforcement Learning können Teile des Reaktionsprozesses automatisieren, was die Reaktionszeiten erheblich verkürzt **Vergleich Machine Learning vs. Deep Learning** sind zwei wesentliche Bereiche des maschinellen Lernens, wobei Deep Learning eine spezifische Unterkategorie von Machine Learning ist. Der entscheidende Unterschied zwischen den beiden liegt in der Fähigkeit von Deep Learning, unstrukturierte Daten durch komplexe künstliche neuronale Netzwerke zu verarbeiten. Dabei werden künstliche neuronale Netze genutzt, um unstrukturierte Informationen wie Texte, Bilder, Töne und Videos in numerische Werte umzuwandeln und zu verarbeiten. Diese Methode ermöglicht es, komplexe Muster und Zusammenhänge in den Daten zu erkennen, ohne dass vorherige manuelle Merkmalsextraktion erforderlich ist. Im Gegensatz dazu ist klassisches Machine Learning in der Regel auf strukturierte Daten angewiesen und erfordert oft manuelles Feature Engineering, um sinnvolle Merkmale aus den Daten zu extrahieren. Ein weiterer Unterschied liegt in der Datensatzgröße und der benötigten Rechenleistung, wo Deep Learning besonders effektiv bei der Verarbeitung dieser großer Datensätze ist, erfordert aber leistungsstarke Computer mit GPUs, um komplexe neuronale Netzwerke zu trainieren. Im Gegensatz dazu kann klassisches Machine Learning oft mit kleineren oder mittelgroßen Datensätzen und einfacherer Hardware arbeiten. Im Unterschied dazu sind die Laufzeiten für das Training von Deep-Learning-Modellen können auch erheblich länger sein als bei klassischen Machine-Learning-Algorithmen, da neuronale Netze eine große Anzahl von Gewichten berechnen müssen. Die Interpretierbarkeit der Ergebnisse ist ein weiterer Unterschiedspunkt: Während einige klassische Machine-Learning-Algorithmen leicht interpretierbare Ergebnisse liefern können, sind die Ergebnisse von Deep Learning oft schwer zu interpretieren. Insgesamt bieten sowohl Deep Learning als auch Machine Learning ein breites Spektrum an Anwendungsmöglichkeiten und können je nach den Anforderungen eines bestimmten Problems oder Projekts eingesetzt werden. Deshalb hängt die Wahl zwischen ML und DL von verschiedenen Faktoren ab, einschließlich der Art der verfügbaren Daten, der gewünschten Genauigkeit und der verfügbaren Rechenressourcen.¹¹⁴

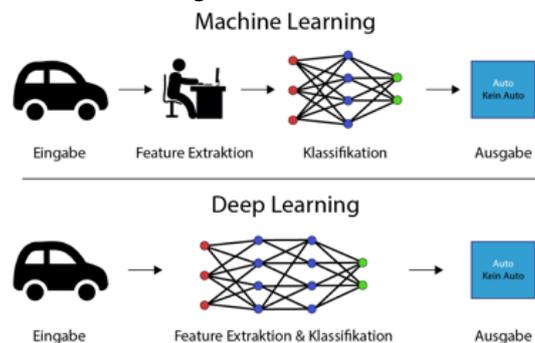


Abbildung 110: Machine Learning vs. Deep Learning: der Unterschied liegt in der Feature Extraktion und dem Einsatz von tiefen, künstlichen neuronalen Netzen

3.1.4. Natural Language Processing

Natural Language Processing (NLP) ermöglicht Computern menschliche Sprache in gesprochener oder geschriebener Form zu verstehen und darauf zu reagieren sowie vereint Aspekte der Sprachwissenschaft, Mathematik und Computertechnologie. Hierbei bietet die NLP zahlreiche Chancen und Möglichkeiten, wie z.B. durch den Einsatz von Artificial General Intelligence (AGI) könnten komplexe Probleme gelöst und neue Technologien entwickelt werden. Weiterhin NLP ermöglicht auch eine sinnvolle Automatisierung von Aufgaben, was zu höherer Effizienz und Produktivität führen kann und dies könnte wiederum zu erheblichen Kosteneinsparungen und einer Verbesserung der Lebensqualität führen, besonders für Menschen mit besonderen Bedürfnissen. Weitere auch Herausforderungen sind im Zusammenhang mit NLP zu betrachten, wie z.B. die Datenschutzbedenken stehen an erster Stelle, da der Zugriff auf persönliche Daten erforderlich ist. Der mögliche Verlust von Arbeitsplätzen durch Automatisierung und der Missbrauch von Daten für die Verbreitung von Falschinformationen sind ebenfalls wichtige Bedenken, dass dazu führt das komplexe NLP-Modelle immer

vgl.¹¹⁴ (Wuttke)

schwerer nachzuvollziehen sind, was zu Kontrollverlust führen kann. Es ist daher entscheidend, NLP-Technologien verantwortungsbewusst einzusetzen und sicherzustellen, dass diese zum Wohle der Gesellschaft genutzt werden und die richtige Balance zwischen Chancen und Herausforderungen kann NLP sein volles Potenzial entfalten und die Art und Weise, wie wir mit Sprache interagieren, grundlegend verändern.¹¹⁵

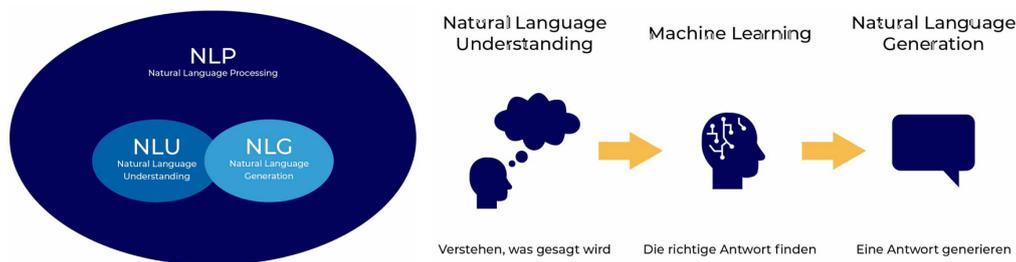


Abbildung 111: Natural Language Processing

Natural Language Understanding (NLU) ist das Verstehen natürlicher Sprache. NLU konzentriert sich auf das maschinelle Verständnis von Sprache, indem es grammatikalische Strukturen und den Kontext analysiert und hilft, Texte zu klassifizieren, Wortarten zu unterscheiden und relevante Informationen aus Inhalten zu extrahieren. NLU wird häufig zur Automatisierung von Aufgaben wie der E-Mail-Kategorisierung und der Textanalyse eingesetzt. **Natural Language Generation (NLG)** ist die Erzeugung natürlicher Sprache. NLG ist das Gegenstück zu NLU und beschäftigt sich mit der automatischen Erstellung von Texten aus Dateneingaben. Typische Anwendungen sind die Generierung von Berichten, Textzusammenfassungen und personalisierten Inhalten und eröffnen neue Geschäftsmöglichkeiten, indem diese große Datenmengen effizient in wertvolle Texte umwandeln.¹¹⁶

NLP	NLU	NLG
NLP ist ein umfassendes Konzept	NLU ist ein spezifisches Konzept	NLG ist ein spezifisches Konzept
NLP trägt neben reinem Textverständnis zur Entscheidungsfindung bei	NLU befasst sich mit dem reinen Textverständnis	NLG befasst sich mit der Generierung von Texten auf Grundlage strukturierter Daten
NLP ist eine Kombination aus NLU und NLG und wird zum Lösen von Problemen mittels künstlicher Intelligenz eingesetzt	NLU ist eine Unterkategorie von NLP	NLG ist eine Unterkategorie von NLP
NLP beinhaltet den gesamten Prozess der Sprachverarbeitung	NLU konzentriert sich auf die Fähigkeit von Textverständnis	NLG sorgt für die Erstellung natürlicher Sprache, sodass diese für den Menschen verständlich ist
NLP wandelt unstrukturierte Daten in strukturierte Daten um	NLU liest Daten und wandelt diese in strukturierte Daten um	NLG erstellt strukturierte Daten

Abbildung 112: Unterschiede von NLP, NLU und NLG

3.2. KI- Lebenszyklus mit Bias

Bei der KI-Anwendungen kann ein Verzerrungseffekt auftreten, bei dem ein sogenanntes Risiko von **Bias** entsteht, d.h. dass Fehler bei der Datenerhebung, ob absichtlich oder unabsichtlich, zu diskriminierenden Entscheidungen führen können. Dies wird behandelt als Unfairness im Algorithmus und die Ursachen von Bias in KI-Systemen liegen oft in vorurteilsbehafteten Trainingsdaten, Auswahl von Modellierungsansätzen und subjektiven Entscheidungen bei der Algorithmengestaltung. Das **Garbage in - Garbage out-Prinzip** (GIGO-Prinzip) zeigt, dass die Qualität der Ausgaben von der Qualität der Eingabedaten abhängt,

vgl.¹¹⁵ (Circle)

vgl.¹¹⁶ (Wuttke, 2023)

dabei wird bei der KI das Ergebnis fehlerhaft oder verzerrt dann ausgegeben (Halluzinationen). Das ML und deren Algorithmen und Modelle sind inbegriffen und arbeiten nur dann perfekt zusammen, wenn die ausgewählten Trainingsdaten mit Verzerrungen trainiert und in ihren Ausgabedaten berücksichtigt werden. Verbesserung der Datenqualität und die Vermeidung von Verzerrungen legen diese technischen und ethischen Grundvoraussetzungen voraus, um die Verantwortung für den Einsatz von KI-Systemen zu garantieren. Dabei müssen Unternehmen Strategien vorher zur Bias-Minimierung integriert werden, wie qualitativ vielfältige Datensätze, transparente Algorithmen, ethische Richtlinien und kontinuierliche Überwachung. Die Förderung von unterschiedlichen zusammengesetzten, fachübergreifenden Teams und externen Audits ist ebenfalls entscheidend, um eine faire Bewertung des Bias in KI-Systemen zu gewährleisten. Insgesamt ist die Bekämpfung von Bias in KI-Systemen bedeutsam, um Systeme zu schaffen, die leistungsfähig, fair und vertrauenswürdig sind. Die Offenlegung der Funktionsweise von KI-Algorithmen hilft, Bias-Quellen aufzudecken und zu korrigieren. Durch die Bereitstellung von Informationen über Daten, Methoden und Entscheidungslogik der Algorithmen sowie die Nutzung von Techniken wie **Local Interpretable Model-Agnostic Explanations (LIME)** und **SHapley Additive exPlanations (SHAP)** können die Einflussfaktoren auf die Entscheidungen der KI zu erkennen. LIME ist ein einfaches lokales Modell um Aussagen zu programmieren und zu analysieren, welche Eingabedatenmerkmale in einem KI-Modell zur Entscheidung führen. SHAP ist eine Spieltheorie mit Fairness, die durchgehende und quantifizierte Vorhersageoptionen beinhaltet. Eine vollständige Offenlegung ist oft nicht möglich, aber selbst teilweise Transparenz kann hilfreich sein, um ein Grundverständnis der Arbeitsweise der KI zu vermitteln. Die Bias besitzen unterschiedliche Arten und Lebenszyklus in der ML Modell Anwendungen. Die KI-Verordnung der EU im Art. 10 Abs. 5 werden spezifische Richtlinien zur Bias-Minimierung gesetzliche Rahmenbedingungen und Vorschriften, die Gleichbehandlung, gerade im Bereich der Identifikation, Überwachung und Korrektur von Verzerrungen im Hochrisiko-KI-Systemen dargelegt.¹¹⁷

Bias kann während des **gesamten Lebenszyklus maschinellen Lernens** entstehen und umfasst verschiedene Phasen, in denen Entscheidungen und Praktiken die Entwicklung und Implementierung von ML-Systemen einwirken und birgt das Potenzial für Bias. Alles beginnt mit der Datenerhebung und es eine Zielgruppe definiert, aus der eine Stichprobe entnommen wird. Merkmale und Kennzeichnungen werden lokalisiert und gemessen, und der Datensatz wird in Trainings- und Testdaten aufgeteilt. Dann folgt das Training eines ML-Modells auf der Grundlage von Trainingsdaten. Die Testdaten dienen dazu, das Modell zu bewerten, danach wird das Modell für den Einsatz in der realen Welt veröffentlicht und trifft Entscheidungen für die Nutzer. Dieser Prozess ist zyklisch: Die Entscheidungen des Modells beeinflussen den Zustand der Welt, der wiederum bei der nächsten Datenerfassung oder Entscheidungsfindung beachtet wird.¹¹⁸ Die wichtigsten Bias-Typen sind: **Datenbias** entstehen, wenn die Trainingsdaten nicht entscheidend für die gesamte Zielpopulation sind, wie z.B. ein Datensatz, der überwiegend Männer enthält, wodurch ein Modell entsteht, das schlechtere Ergebnisse für Frauen liefert; **Algorithmischer Bias** sind Daten neutral, kann ein Algorithmus durch seine Struktur oder Logik Verzerrungen entwickeln, die zu logischen Fehlern führen; **Menschlicher Bias** durch Entwickler subjektiv ihre eigenen Vorurteile in den Prozess einbringen, was sich auf die Auswahl der Merkmale, die Definition der Ziele oder die Interpretation der Ergebnisse widerspiegeln kann; **Interaktionsbias** treten auf, wenn Nutzer durch ihre Interaktionen mit dem System unwissend eine Verzerrung erzeugen, indem sie bestimmte Muster bevorzugen oder wiederholen.¹¹⁹

Bias in KI-Systemen kann zu unfairer oder diskriminierender Entscheidungsfindung führen, daher gibt es verschiedene Strategien, von der Formulierung von Anwendungen, um Bias weitestmöglich zu minimieren, über eine neutrale Datenerfassung bis hin zur Entwicklung von Algorithmen zur Verringerung von Voreingenommenheit. Dabei können unfaire Entscheidungen auftreten. Hierzu gibt es zwei Ansätze zur Bekämpfung, wie die Förderung von Diversität in Teams sollten unterschiedlich zusammengesetzt sein, um unterschiedliche Perspektiven einzubeziehen und Bias auszuschließen sowie die erklärbare KI, die KI-Systeme

vgl.¹¹⁷ (Paschou, 2024)

vgl.¹¹⁸ (Ulm, 2023)

vgl.¹¹⁹ (FAIRNESS, 2024)

nachvollziehbarer, damit Menschen sie besser verstehen und überprüfen können, hinsichtlich in Bezug auf Bias und Fairness.¹²⁰

Im nachfolgenden werden die Techniken zur Minimierung von Bias in Algorithmen erklärt, die in drei Kategorien eingeteilt werden können:¹²¹

Präprozessing-Algorithmen (Pre-Processing Algorithmen) liegen die Diskriminierung in den Daten vor der eigentlichen Modellverarbeitung, die zu minimieren sind. Diese Algorithmen stützen sich auf einem binären oder spezifischen sensiblen Attribut und passen die Verteilungen der Merkmale so an, dass diese für **jede sensible Gruppe gleich** sind. Das Ergebnis ist ein veränderter Datensatz, bei dem die Merkmale vom sensiblen Attribut abgespalten sind. Dadurch soll verhindert werden, dass ein auf diesen Daten trainiertes Modell Unterscheidungen auf Grundlage der sensiblen Attribute erlernt.

In-Prozessing-Algorithmen abändern den Trainingsprozess moderner Lernverfahren, um Diskriminierung zu minimieren. Dies kann durch Anpassungen der Zielfunktion oder die Implementierung von Einschränkungen zur Reduzierung von Bias entstehen. Ein Regularisierungsterm wird benutzt, um die gegenseitige Abhängigkeit zwischen den Vorhersagen und den sensiblen Attributen herabzusetzen. Dieser Term wird in das Optimierungsziel eingefügt, sodass durch die Reduzierung der Zielfunktion sowohl eine genaue Vorhersage als auch die Vermeidung einer übermäßigen Abhängigkeit von den sensiblen Attributen erreicht wird und somit eine **demografische Gleichheit** gefördert wird.

Post-Prozessing-Algorithmen bezeichnet das **Modell als Blackbox**, sodass weder die Trainingsdaten noch der Lernalgorithmus nicht verändert werden können. Diese Algorithmen ordnen die ursprünglichen Modellvorhersagen nachträglich durch eine Funktion neu zu. Die Fairness-Definition „**Equalized Odds**“ verlangt, dass sowohl die **True Positive Rate** (TPR) als auch die **False Positive Rate** (FPR) für jede sensible Gruppe gleich sind, während bei „**Equal Opportunity**“ **nur die TPR gleich** sein muss. Um alle diesen Bedingungen zu nachzukommen, passen Post-Prozessing-Algorithmen die Entscheidungsschwellen für jede Gruppe an, die für die Vorhersage genutzt werden. Da diese Algorithmen nur Zugriff auf den Modell-Output und die sensiblen Attribute benötigen, sind diese sehr flexibel verwendbar und bieten eine gerechtere Entscheidungsfindung während des Modellierungsprozesses.¹²²

Laut **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)** werden acht Bias Arten definiert:¹²³

Attrition-Bias tritt auf, wenn Studienteilnehmer die Studie vorzeitig abbrechen, was zu einer Verzerrung der Ergebnisse führen kann und eine Intention-to-Treat-Analyse kann diesen Bias minimieren.

Detection-Bias entsteht durch unterschiedliche Methoden zur Feststellung von Endpunkten in Studiengruppen, was die Genauigkeit der Ergebnisse verfälscht und entgegenwirkt durch eine Verwendung einheitlicher Untersuchungsverfahren und verblindeter Erhebungen.

Performance-Bias besteht, wenn eine Gruppe zusätzliche Behandlungen erhält, die nicht untersucht werden und Verblindung des ärztlichen Personals kann Abweichungen in der Behandlung minimieren.

Publication-Bias führt dazu, dass Studien mit nicht positiven Ergebnissen seltener veröffentlicht werden, was zu einer Überschätzung des Effekts führen kann und die Einbeziehung unveröffentlichter Studien in Meta-Analysen kann diesen Bias minimieren.

Selektionsbias entsteht durch unterschiedliche Zusammensetzungen der Vergleichsgruppen und durch eine zufällige Zuteilung von Teilnehmern kann diese Verzerrung reduziert werden.

Lead-Time-Bias beeinflusst die Beurteilung von Früherkennungsmethoden und kontrollierte Studien können diesen Bias minimieren, indem alle Teilnehmer zu einem einheitlichen Zeitpunkt beobachtet werden.

vgl.¹²⁰ (Ulm, 2023)

vgl.¹²¹ (Ulm, 2023)

vgl.¹²² (Ulm, 2023)

vgl.¹²³ (IQWiG)

Length-Bias entsteht, wenn Früherkennungsmethoden Erkrankungen mit langsamem Verlauf bevorzugen und kontrollierte Studien mit einer Gruppe, der Früherkennung angeboten wird und einer Gruppe ohne, können diesen Bias reduzieren.¹²⁴

Historische Verzerrungen beziehen sich auf bereits existierende Vorurteile und soziotechnische Probleme, die in den Datenerstellungsprozess hineinfließen können – selbst bei einer idealen Stichprobenziehung und Merkmalsauswahl.

Repräsentationsverzerrungen entstehen durch die Art der Datenerhebung, wobei nicht repräsentative Stichproben die Unterschiedlichkeit der zugrunde liegenden Population unzureichend darstellen.

Messfehler treten auf, wenn Merkmale und Kennzeichnungen, die für ein Vorhersageproblem ausgewählt, erfasst oder berechnet werden, lediglich als Proxies arbeiten und ein nicht direkt kodierbares oder beobachtbares Konzept annäherungsweise erstellt wird.

Omitted Variable Bias entsteht, wenn wichtige Variablen aus dem Modell ausgeschlossen werden, was zu falschen oder verzerrten Vorhersagen führen kann.

Evaluation Bias besteht, wenn die Trainingsdaten die tatsächliche Nutzerpopulation nicht darstellen oder die gewählten Leistungsmetriken bestehende Verzerrungen weiter verstärken.

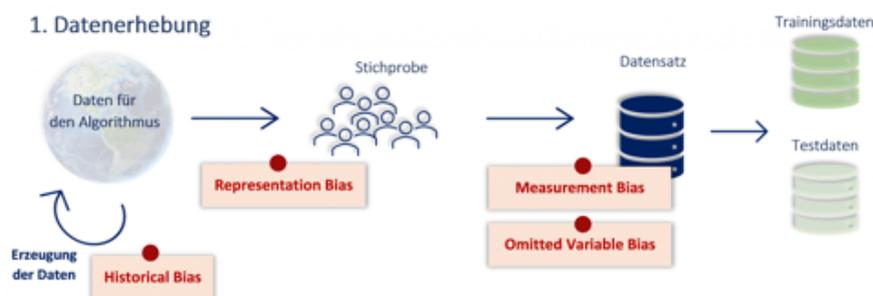


Abbildung 113: Datenerhebung von Bias

Algorithmische Verzerrungen entstehen direkt durch den Algorithmus selbst und sind unabhängig von den zugrunde liegenden Daten.

Aggregationsverzerrungen entstehen, wenn ein einzelnes Modell auf Daten angewendet wird, die eigentlich unterschiedlich behandelt werden sollten.

User Interaction Bias bezieht sich auf Verzerrungen, die durch die Benutzeroberfläche oder das Nutzerverhalten entstehen und einseitige Interaktionen fördern können.

Population Bias entsteht bei Statistiken, demografischen Daten oder Nutzereigenschaften auf einer Plattform, die von denen der ursprünglichen Zielgruppe abweichen.

Deployment Bias bestehen, wenn ein System während seines Einsatzes falsch verwendet oder interpretiert wird und so zu Fehlentscheidungen führen und **Feedbackschleifen** zwischen Daten, Algorithmen und Nutzern bestehende Verzerrungen verstärkt und perpetuiert werden.¹²⁵

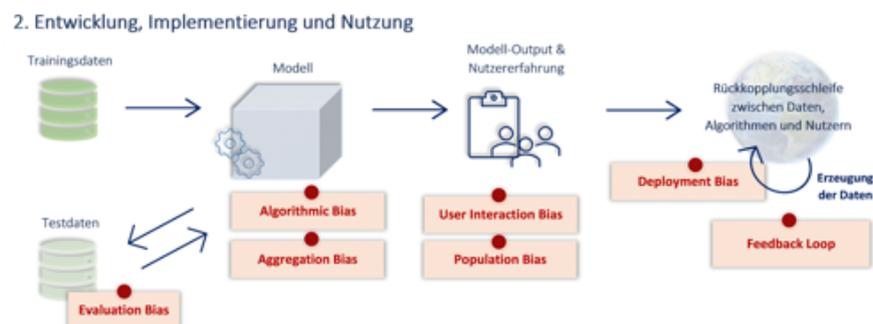


Abbildung 114: Entwicklung, Implementierung und Nutzung von Bias

Bias kann in jedem Schritt des **ML-Lernzyklus** auftreten, der die Phasen der Datenerfassung, Modellentwicklung und Systembereitstellung umfasst. Zu Beginn der Datener-

vgl.¹²⁵ ((IBA), 2023)

fassung wird eine Zielpopulation festgelegt und daraus eine Stichprobe entnommen, dabei werden Merkmale und Attribute identifiziert, gemessen und der Datensatz anschließend in Trainings- und Testdaten unterteilt. Auf Basis der Trainingsdaten wird ein ML-Modell entwickelt, welches anschließend auf die Testdaten angewendet wird. Nach erfolgreicher Modellentwicklung erfolgt die Implementierung in einer realen Umgebung, in der das Modell Entscheidungen für Nutzer trifft. Da dieser Prozess zyklisch ist, besteht in jeder Phase die Möglichkeit, verschiedene Arten von Verzerrungen zu integrieren, die sich auf die Modellleistung und -fairness auswirken können.^{126 127}

Im folgenden Abschnitt wird der formelle Aufbau einer KI-Anwendung, mit Hilfe der Ausführungen des Fraunhofer Instituts, betrachtet:¹²⁸

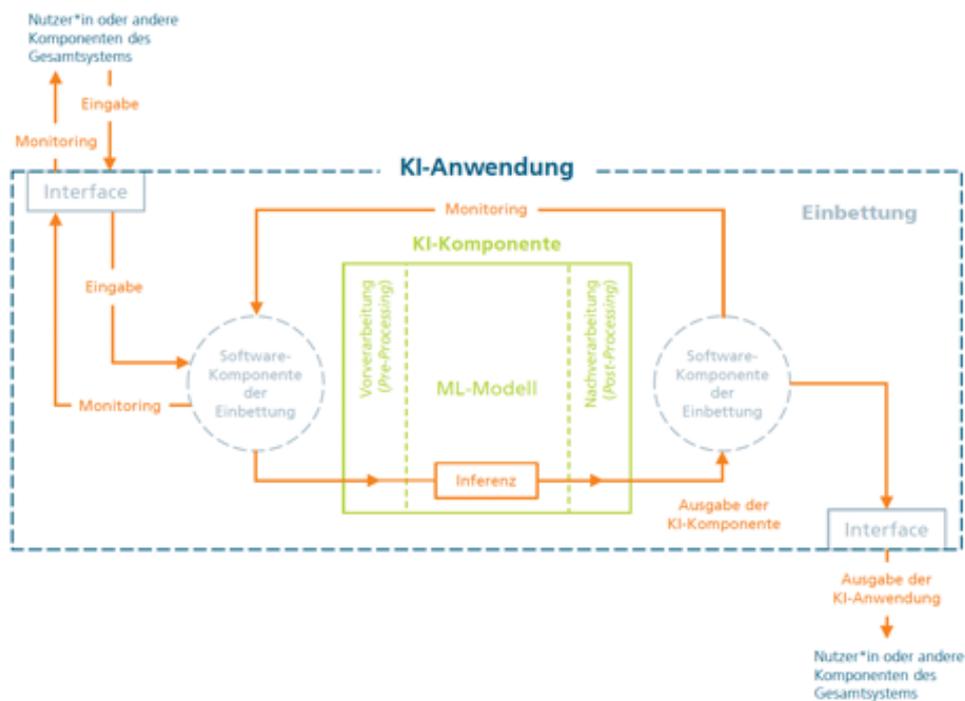


Abbildung 115: Formaler Aufbau einer KI-Anwendung

Die Künstliche Intelligenz (KI) konzentriert sich auf KI-Anwendungen, in Zusammenhang mit Machine Learning (ML) und Deep Learning (DL), die mit neuronalen Netzen umgesetzt werden, aber auch andere Methoden wie Entscheidungsbäume oder Stützvektorverfahren, je nach Entwicklung einer Anwendung, um Anforderungen hinsichtlich Transparenz oder Sicherheit zu erfüllen, müssen betrachtet werden. Das ML-Modell bildet den Kern der KI-Komponente und wird durch Vor- und Nachbereitungsschritte für die Ein- und Ausgaben des ML-Modells ergänzt und auch die KI-Komponente sind mathematische Objekte, die als Grundlage für ein Funktionsprinzip Eingabe-Ausgabe-Mappings dienen, wie z.B. Objekterkennung, Protokollierung / Überwachung von verschiedenen Verarbeitungsschritten bei den Eingabedaten. Die klare Definition der Grenzen einer KI-Anwendung innerhalb eines Gesamtsystems ist massgeblich, wenn das Gesamtsystem nicht ausschließlich auf KI-Technologien basiert. Diese Definition richtet sich nach der spezifischen Funktionalität der KI-Anwendung im Kontext des Gesamtsystems, wie eine auf ML basierende Objekterkennung in verschiedene Systeme implementiert sein können:

Bedrohungserkennung und -abwehr werden genutzt, um verdächtige Aktivitäten in IT-Systemen zu erkennen, potenzielle Bedrohungen, die auf Anomalien hinweisen, um Sicherheitsverletzungen zu verhindern oder zu bekämpfen.

vgl.¹²⁶ (Paschou, 2024)

vgl.¹²⁷ (FAIRNESS, 2024)

vgl.¹²⁸ (Dr. Maximilian Poretschkin, et al.)

Verhaltensanalyse wird eingesetzt, um Analyse von Benutzeraktivitäten und Systemereignissen auf ungewöhnliche Verhaltensmuster, die auf Sicherheitsrisiken oder Insider-Bedrohungen hinweisen könnten.

Automatisierung von Sicherheitsvorgängen werden verwendet, um Sicherheitsprozesse zu automatisieren, wie die Erkennung und Entfernung von Malware, die Anpassung von Zugriffsrechten oder die Reaktion auf Sicherheitsvorfälle in Echtzeit.

Bedrohungsinformationen und -analyse werden in Systemen implementiert, um große Mengen an Sicherheitsdaten zu analysieren und Muster zu erkennen, die auf neue Bedrohungen oder Angriffstechniken hinweisen könnten, und um Sicherheitsanalysten bei der Bewertung und Priorisierung von Sicherheitsereignissen zu unterstützen.

Vorausschauende Sicherheit durch die Analyse historischer Daten und das Erkennen von Trends kann KI beitragen, zukünftige Sicherheitsbedrohungen vorherzusagen und proaktiv Maßnahmen zu ergreifen, um Sicherheitslücken zu schließen oder Schwachstellen zu beheben, bevor sie ausgenutzt werden können.

Adaptive Sicherheitsmaßnahmen werden entwickelt, um Sicherheitsmaßnahmen dynamisch anzupassen und zu optimieren, die auf veränderte Bedrohungsszenarien, Benutzeraktivitäten oder Systemkonfigurationen hinweisen.

Identitäts- und Zugriffsmanagement werden konfiguriert, um die Identität und den Zugriff von Benutzern zu überwachen und zu verwalten, indem diese verdächtige Aktivitäten oder ungewöhnliche Zugriffsversuche erkennt und angemessene Maßnahmen ergreift, um das Risiko von Datenverstößen zu minimieren.

Compliance-Überwachung wird durchgeführt, um kontinuierliche Überwachung von Sicherheitsrichtlinien und -vorschriften gegen Compliance-Verstöße zu lokalisieren und zu beheben, um die Einhaltung gesetzlicher Anforderungen sicherzustellen.

In solchen Szenarien werden Softwaremodule eingesetzt, die die Konsistenz der Ausgaben der KI-Komponente überprüfen, als integraler Bestandteil der KI-Anwendung betrachtet während andere Komponenten nicht dazugehören.

Das **ML-Modell**, wie schon beschrieben, ist ein mathematisches, abstraktes Objekt, das mithilfe eines maschinellen Lernverfahrens erstellt wurde, um problematische Eingabe-Ausgabe-Relation zu lösen. Dabei besteht das **Modell Neuronaler Netze** aus einer Auflistung an Hyperparametern, gelernten Parametern sowie einer Beschreibung ihrer Interaktion zur Laufzeit (Architektur) und liefert somit die funktionelle Grundlage der KI-Anwendung. Das Modell liefert die Grundlage für die Durchführung einer Klassifikationsaufgabe, wobei die Eingabe den zu klassifizierenden Gegenstand darstellt und seine Aussage über seine Klasse trifft. In einigen Fällen (bei generativen Modellen) kann die Eingabe (Zufallszahlen) für die tatsächliche Funktionsweise von untergeordneter Bedeutung sein.

Die **KI-Komponente** wird als mathematisches Objekt dargestellt und setzt sich aus dem ML-Modell sowie spezifischen Verfahren zur Datenvorverarbeitung und zur Nachverarbeitung der Modell-Ausgaben zusammen.

Die **Einbettung** definiert die Gesamtheit der umgebenden Komponenten, die unmittelbar auf die Funktionsweise der KI-Komponente eingehen, mit anderen Software-Modulen und technischen Komponenten, die für die Speicherung von Daten oder die Umsetzung von physischen Reaktionen auf die Ausgaben der KI-Komponente dienen. Die Einbettung umfasst alle umgebenden Komponenten, die sich direkt auf die Funktionalität und den Betrieb der KI-Komponente auswirken. Hierzu gehören Software-Module, die die KI-Komponente aktivieren, ihre Ergebnisse verarbeiten und eine Interaktion ermöglichen und darüber hinaus können Einbettungssoftware-Module verwendet werden, um ein Versagen der KI-Komponente zu erkennen und zu korrigieren.

Interface (Schnittstelle) ist ein Teil der Einbettung, der die Interaktion der KI-Anwendung mit der Außenwelt ermöglicht und bietet verschiedene Interaktionsmöglichkeiten für Funktionalität und Design einer KI-Anwendung, die durch Input von Daten generiert werden, die Ergebnisse / Ausgaben der KI-Komponente nach außen ausgegeben bzw. abrufbar gemacht werden.

Die **KI-Anwendung** selbst steht für das Eingabe-Ausgabe-Mapping in einem gegebenen Einsatzkontext, das auf der implementierten KI-Komponente basiert. Der Prüfkatalog betrachtet nicht isoliert das ML-Modell oder die KI-Komponente, sondern untersucht die Funk-

tionalität des Eingabe-Ausgabe-Mappings in einem gegebenen Einsatzkontext. Dabei wird bewertet, ob die Verarbeitungsschritte und Ergebnisse der KI-Komponente im Zusammenspiel mit anderen Komponenten angemessen und frei von Fehlern, Diskriminierung sowie sicher vor Angriffen sind. Dabei liegt der Fokus nicht auf den zugrunde liegenden mathematischen Konzepten, sondern auf der Funktionalität des Eingabe-Ausgabe-Mappings. Eine KI-Anwendung kann eigenständig oder in ein Gesamtsystem integriert sein.

Der **Lebenszyklus einer KI-Anwendung** dient zur Identifizierung und Reduzierung von KI-Risiken und sollte daher bei der Bewertung der Qualität umfassend berücksichtigt werden, besonders bei den Daten, die die Funktionalität der Anwendung oft direkt aus diesen ableiten. Dabei stehen die Prozesse des Data Mining im Vordergrund, die eng mit der Entwicklung und dem Betrieb klassischer IT-Systeme verbunden sind. Der CRISP-DM-Standard, der den Data-Mining-Prozess von der Zielsetzung über die Datenauswahl bis hin zur Anwendungsentwicklung und dem Betrieb betrachtet, kann auch auf die komplexeren Modellklassen von KI-Anwendungen wie tief neuronale Netze übertragen werden.

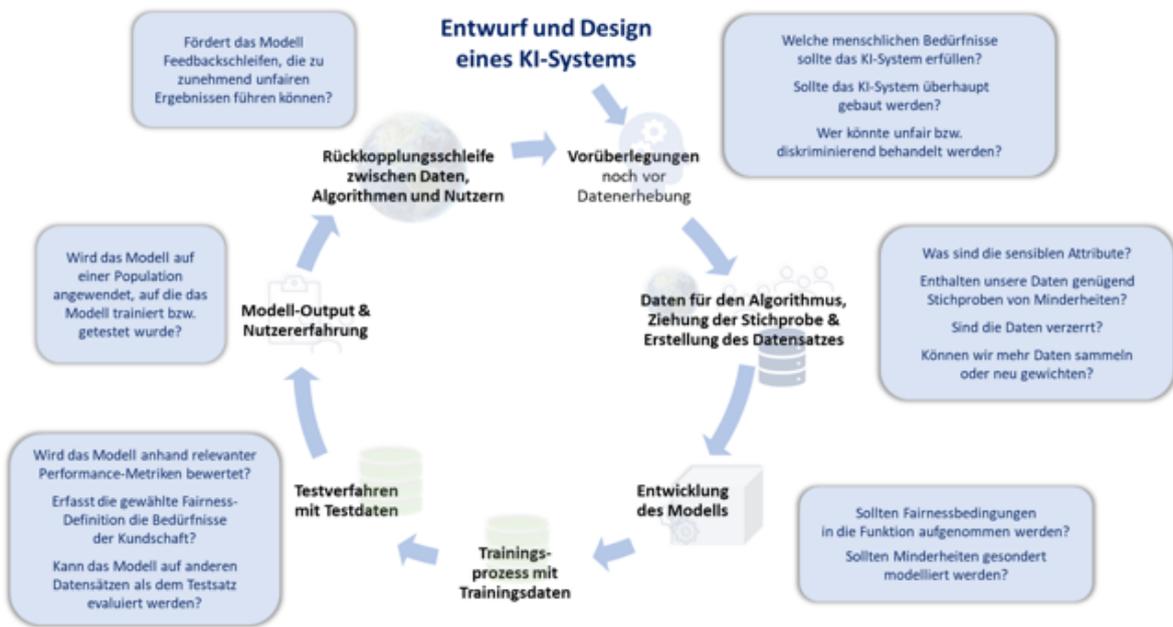


Abbildung 116: Lebenszyklus einer KI-Anwendung

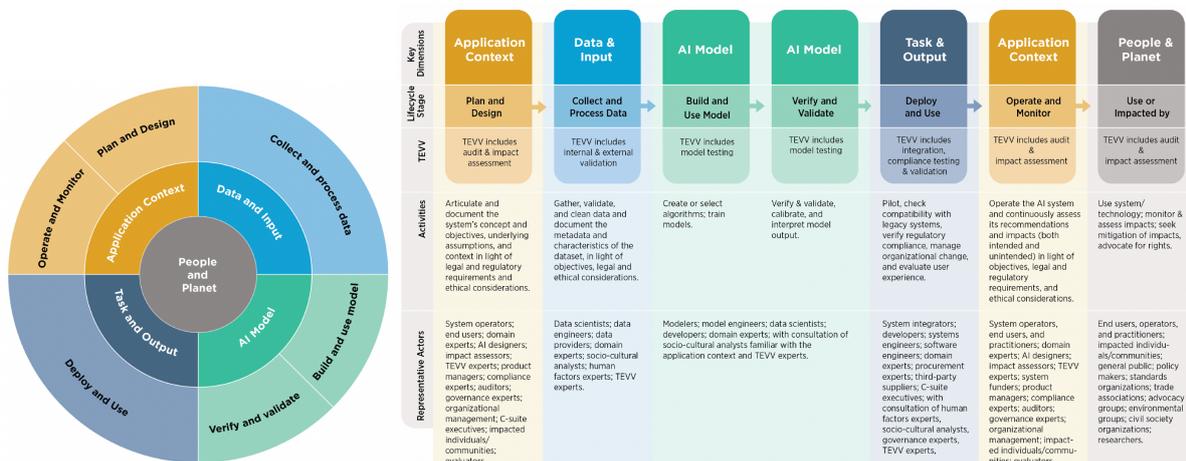


Abbildung 117: Abstrahierter Lebenszyklus einer KI-Anwendung

In der **ersten Phase** wird festgelegt, in welchem Kontext das SIEM-System mit KI eingesetzt werden soll und welche Anforderungen es erfüllen muss. Es wird definiert, welche Sicherheitsbedrohungen erkannt und welche Daten überwacht werden sollen. Zunächst werden die Anwendungsbereiche identifiziert, indem spezifische Bedrohungen, wie Netzwerksicherheit, Datenlecks oder Insider-Bedrohungen, festgelegt werden. Anschließend werden die Ziele des KI-gesteuerten SIEM-Systems definiert, z. B. die automatisierte Erkennung von Sicherheitsvorfällen oder die Reaktionsautomatisierung. Stakeholder wie Sicherheitsverantwortliche, IT-Administratoren oder Compliance-Officer müssen ebenfalls identifiziert werden. Zudem werden regulatorische Anforderungen, wie die Einhaltung von **GDPR** oder **ISO 27001**, berücksichtigt. Fehlende oder unvollständige Anforderungen könnten zu einem ineffektiven SIEM-System führen, weshalb detaillierte Anforderungsanalysen notwendig sind.

In der **zweiten Phase** geht es um die Datenakquise und -vorbereitung. Der Datenfluss des SIEM-Systems wird eingerichtet, indem verschiedene Datenquellen erkannt und implementiert werden, z. B. Netzwerklaufwerke, Firewalls oder Benutzeraktivitäten. Eine sorgfältige Datenbereinigung ist erforderlich, um Anomalien zu vermeiden, die durch nicht stimmige oder fehlerhafte Daten verursacht werden könnten. Zudem werden Metadaten und Indikatoren festgelegt, um Bedrohungen oder anomales Verhalten zu erkennen. Unvollständige oder ungenaue Daten können zu Fehlalarmen oder übersehenen Bedrohungen führen, daher ist eine sorgfältige Datenintegration und -bereinigung essenziell.

Die **dritte Phase** widmet sich der Modellierung und Entwicklung von KI-Algorithmen. Hierbei werden Modelle des maschinellen Lernens entwickelt, um sicherheitsrelevante Muster zu erkennen sowie umfasst die Entwicklung von Algorithmen zur Anomalieerkennung und die Korrelation von Ereignissen aus verschiedenen Datenquellen. Das Modell wird mithilfe historischer Bedrohungsdaten trainiert und getestet, um sicherzustellen, dass neue Bedrohungen zuverlässig erkannt werden. Ein schlecht trainiertes Modell könnte jedoch zu Fehlalarmen oder übersehenen Bedrohungen führen und um dies zu vermeiden, werden umfangreiche historische Daten verwendet und regelmäßige Modellanpassungen durchgeführt.

In der **vierten Phase** erfolgt die Bereitstellung und Integration des KI-gesteuerten SIEM-Systems in die bestehende Sicherheitsinfrastruktur. Dabei wird darauf geachtet, dass das System mit bestehenden IT-Sicherheits- und Netzwerkumgebungen kompatibel ist. Ein Testbetrieb wird durchgeführt, um sicherzustellen, dass das System korrekt arbeitet und keine unnötigen Fehlalarme auslöst. Nichtüberbeinstimmungen oder falsch konfigurierte Alarme können zu Sicherheitslücken führen, deshalb gründliche Tests und regelmäßige Überprüfungen der Systemkompatibilität wichtig sind.

In der **fünften Phase** wird das SIEM-System in Betrieb genommen und kontinuierlich überwacht und dabei das KI-SIEM-System in Echtzeit eingehende Datenströme analysiert, potenzielle Bedrohungen identifiziert und diese durch das Incident-Response-Team eskaliert. In bestimmten Fällen kann das System automatisierte Reaktionen, wie das Sperren eines kompromittierten Benutzers, einleiten. Das Sicherheitsteam bestätigte die Alarme und entschied über das weitere Vorgehen. Fehlalarme oder übersehene Bedrohungen könnten das Vertrauen in das System untergraben, weshalb Feedbackschleifen integriert werden, um das System kontinuierlich zu verbessern.

Die **sechste Phase** konzentriert sich auf das Testen, Evaluieren, Verifizieren und Validieren (TEVV) des KI-SIEM-Systems. Regelmäßige Tests werden durchgeführt, um sicherzustellen, dass das System neuen Bedrohungen gewachsen ist. Die vom System erzeugten Alarme werden getestet, um ihre Relevanz für sicherheitskritische Vorfälle zu überprüfen. Die zugrunde liegenden Testergebnisse werden in den Modellen laufend angepasst und optimiert, um Erkennungsraten zu verbessern und Fehlalarme zu reduzieren. Unzureichende Tests könnten dazu führen, dass das SIEM-System neuen Bedrohungen nicht standhält, weshalb regelmäßige Sicherheitsprüfungen und Modellüberwachung implementiert werden.

In der **siebten Phase** stehen die Governance und ethische Kontrolle im Vordergrund. Die Nutzung von KI im Sicherheitsbereich muss ethisch und rechtlich vertretbar sein, insbesondere im Hinblick auf Datenschutz und Überwachung. Es muss sichergestellt werden, dass alle erfassten Daten im Einklang mit Datenschutzgesetzen verarbeitet werden. Aber auch die Nachvollziehbarkeit der KI-Entscheidungen muss gewährleistet sein, die durch regelmäßige Audits und Dokumentationen sicherstellen, dass das KI-SIEM-System den Sicher-

heitsvorgaben entspricht. Fehlende Governance könnte zu Verstößen gegen Datenschutzrichtlinien führen oder unethische Überwachungsmaßnahmen fördern.

Der KI-gesteuerte Lebenszyklus eines SIEM-Systems sollte in allen Phasen von der Anforderungenanalyse bis hin zur kontinuierlichen Überwachung und Optimierung optimiert entwickelt werden, da jede Phase entscheidend ist, um die Sicherheit der IT-Infrastruktur zu gewährleisten und Bedrohungen frühzeitig zu erkennen. Eine gründliche Risikobewertung und kontinuierliche Verbesserung des Systems tragen dazu bei, Fehlalarme zu reduzieren und die Sicherheit zu maximieren.¹²⁹

Die Auswahl und Aufbereitung von Daten sind entscheidend für die Funktionalität einer KI-Anwendung, da Eigenschaften wie die gelernten Gewichte direkt aus der Datenaufbereitung resultieren. Im Vergleich zum traditionellen Data Mining übernehmen komplexe maschinelle Lernverfahren oft das manuelle Vorverarbeiten von Daten, z.B. das Herausarbeiten relevanter Features und daher müssen Datenrisiken in der Qualitätsbewertung stärker berücksichtigt und angemessene Maßnahmen ergriffen werden, z.B. zur Fairness oder zur Qualität von Labels.

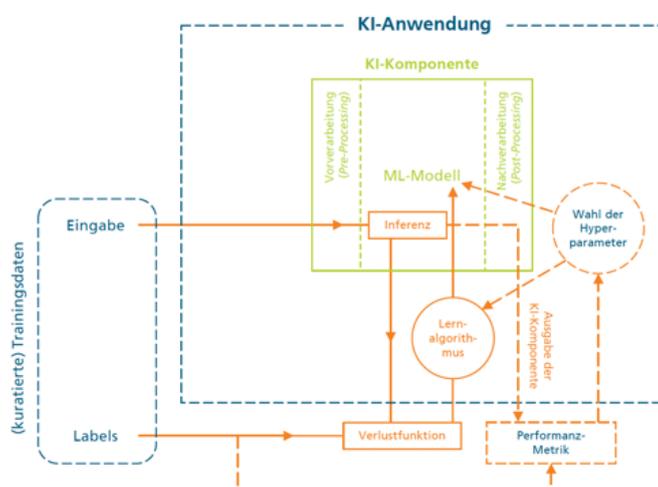


Abbildung 118: Training des ML-Modells einer KI-Anwendung

Die **Entwicklung einer KI-Anwendung** durchlaufen verschiedene Bereiche im Design, Modellbildung und Testing und beinhalten KI-Komponente, wie Designentscheidungen, Auswahl der Modellarchitektur, Einbettungskomponenten sowie Implementierung von Konsistenzprüfungen, Implementierung und Ausfallredundanz, um Systemsqualität zu gewährleisten.

Das **Training des maschinellen Lernmodells** erfordert eine präzise Koordination der verschiedenen Hyperparameterauswahl, Techniken des ML wie z.B. die Initialisierung der Gewichte auf Trainingsdaten, Analyse der erzeugten Ergebnisse von ML-Verfahren, mit den Daten der entsprechenden **Ground Truth** und einer Verlustfunktion zu wägen. Durch eine wiederholende Gewichtsanpassung optimiert, welche einen grossen Einfluss auf das finale Ergebnis hat. Dabei sind verschiedene Qualitätskriterienbewertungen des ML-Modells sehr wichtig und deshalb werden Funktionen, wie Hyperparameterkonfigurationen und separaten Validierungssätzen entgegengestellt, optimale Konfigurationsprobleme herausgefiltert, Verlustfunktionoptimierung durchgeführt, Leistungsmetriken durch Testing überprüft, um sicherzustellen, dass die KI-Anwendung unter normalen Betriebsbedingungen zuverlässig funktioniert und potenzielle Schwachstellen im Modell erkannt werden. Hier können **Eingriffs- und Feedbackmöglichkeiten** integriert werden und somit Risiken mit **Concept- und Model-Drift** zu erkennen. **Concept Drift** sind Eingabedatenveränderungen, hingegen **Model Drift** die Betriebsanpassung im ML-Modells neu zu bewerten, um Anforderungenmanagement durchzuführen und kontinuierlich die KI-Anwendungsverhalten zu überwachen und Zuverlässigkeit sicherzustellen.¹³⁰

vgl.¹²⁹ (NIST, 2023)

vgl.¹³⁰ (Dr. Maximilian Poretschkin, et al.)

3.3. KI Ethik und Herausforderungen

Die **Integration von künstlicher Intelligenz (KI)** in Überwachungssystemen unterliegen ethische Fragestellungen und komplexe Herausforderungen, wie z.B. zentraler ethischer Schutz der Privatsphäre. Überwachungstechnologien in persönliche Lebensbereiche von Einzelpersonen einzugreifen und angemessene Datenschutzmaßnahmen zu integrieren, um die Vertraulichkeit persönlicher Informationen zu gewährleisten.

Ein weiterer **ethischer Grundsatz** betrifft die Transparenz und Verantwortlichkeit von Überwachungssystemen, um Vertrauen in der Öffentlichkeit zu gewinnen, müssen diese Technologien transparent agieren und klare Regelungen müssen die Verantwortlichkeit für ihren Einsatz regeln. Dies schließt auch die Notwendigkeit ein, sicherzustellen, dass Überwachungssysteme, um nicht nur effektiv, sondern auch ethisch verantwortungsbewusst arbeiten.

Die **Gefahr von Diskriminierung und Ungerechtigkeiten durch KI-gesteuerte Überwachungssysteme** durch KI-Algorithmen kann bestehende Vorurteile verstärken, wodurch bestimmte Gruppen bevorzugt oder benachteiligt werden. Daher ist es unabdingbar, Überwachungssysteme auf ihre Fairness zu überprüfen, um jegliche Form von Diskriminierung zu vermeiden.

Ein weiterer **ethischer Leitfaden** betrifft die Zweckbindung von Überwachungstechnologien, die sich auf die ursprünglich vorgesehenen Zwecke beschränken, um Missbrauch für unethische oder nicht genehmigte Zwecke zu verhindern.

Die **Herausforderungen** in diesem Zusammenhang sind ebenfalls vielfältig an Mangel an einheitlichen Standards und angemessenen rechtlichen Rahmenbedingungen für den Einsatz von KI im Überwachungsbereich. Ebenso gestaltet sich die Kontrolle über die komplexen KI-Systeme schwierig, was die Entwicklung von Mechanismen zur Vorhersagbarkeit und Kontrolle erforderlich sind. Die Qualität der Eingangsdaten, die von Überwachungssystemen verwendet werden, beeinflusst maßgeblich deren Leistung und sollte durch eine sorgfältige Überprüfung der Datenqualität durchgeführt werden und somit der Vermeidung von Verzerrungen und Voreingenommenheit zu entgegen.

Schließlich besteht das Risiko des **Missbrauchs von Überwachungstechnologien** für autoritäre oder unethische Zwecke. Daher sind klare Schutzmechanismen und Überwachungsmechanismen erforderlich, um sicherzustellen, dass der Einsatz von KI im Überwachungsbereich im Einklang mit grundlegenden Werten und Menschenrechten steht.

Insgesamt erfordert die **ethische Entwicklung und Anwendung von KI-Überwachungstechnologien** eine ganzheitliche Herangehensweise, bei der Ethikexperten, Juristen, Regierungen und die Gesellschaft als Ganzes aktiv zusammenarbeiten müssen, um angemessene Richtlinien und Kontrollmechanismen zu etablieren. Nur so kann gewährleistet werden, dass diese Technologien nicht nur effektiv, sondern auch ethisch vertretbar eingesetzt werden.¹³¹

3.4. KI Use Case

Die **Überwachung von Künstlicher Intelligenz (KI)-Anwendungen** spielt eine entscheidende Rolle beim Monitoring von IT-Systemen, um deren Leistung, Sicherheit und Genauigkeit zu gewährleisten. Dieser Anwendungsfall konzentriert sich auf die fortlaufende Überwachung von KI-Anwendungen über ihren gesamten Lebenszyklus, angefangen von der Entwicklung bis hin zur Bereitstellung im produktiven Betrieb.

Die **Leistungsüberwachung** ist von entscheidender Bedeutung, um sicherzustellen, dass die KI-Anwendung effizient arbeitet und die gewünschten Ergebnisse in akzeptabler Zeit liefert. Hierbei liegt der Fokus auf der Optimierung von KI-Modellen und Algorithmen.

Die **Genauigkeitsbewertung** beinhaltet die kontinuierliche Überprüfung der Vorhersagegenauigkeit und die Anpassung der Modelle, um sicherzustellen, dass sie den sich ändernden Anforderungen gerecht werden.

vgl.¹³¹ (BSI)

Durch **Echtzeitüberwachung** erfolgt eine schnellere Erkennung und Behebung von Fehlern oder Anomalien in Echtzeit, um eine schnelle Reaktion zu ermöglichen und Ausfallzeiten zu minimieren.

Automatisierte Warnungen werden eingerichtet, um bei Abweichungen von definierten Schwellenwerten automatisch Alarme auszulösen.

Die **Integration relevanter Datenquellen**, einschließlich Trainingsdaten, Echtzeitdaten und Modellmetriken, ermöglichen eine umfassende Überwachung der KI-Anwendung.

Die **Sicherheitsüberwachung** umfasst die Implementierung von Funktionen zur Identifizierung möglicher Bedrohungen oder Angriffe auf die KI-Anwendung.

Der **adaptive Lernprozess** gewährleistet die kontinuierliche Anpassung der Modelle basierend auf Echtzeitdaten und Leistungsrückmeldungen.

Die **Sicherheit und Compliance** werden durch die Überwachung von Sicherheitsbedrohungen und die Gewährleistung der Einhaltung von Datenschutz- und Compliance-Anforderungen sichergestellt.

Das **Monitoring von Künstlicher Intelligenz (KI)** im Kontext von **Cyberangriffen auf Netzwerksysteme** ist ein entscheidendes Hilfsmittel bei der proaktiven Erkennung, Analyse und Abwehr von potenziellen Bedrohungen. Dieser Anwendungsfall fokussiert sich auf die Integration von KI in die Netzwerküberwachung, um Anomalien und mögliche Cyberangriffe zu identifizieren.

Die **Integration von Daten aus unterschiedlichen Netzwerkquellen**, darunter Protokolle, Datenpakete und Endpunkte bilden die Grundlage für eine umfassende Analyse.

Künstliche Intelligenz-Algorithmen durch den Einsatz von Machine-Learning- und KI-Algorithmen erfolgt die Analyse des Netzwerkverhaltens zur frühzeitigen Identifikation von Anomalien.

Die Einbindung von **Bedrohungsinformationen** aus externen Quellen ermöglicht die Erkennung aktueller Angriffsmuster.

Schnellere Echtzeiterkennung von Nutzung der Echtzeitinformationen ermöglicht eine prompte Reaktion auf identifizierte Bedrohungen, Anomalien und die Implementierung von Gegenmaßnahmen zur Minimierung von Cyber-Angriffseinflüssen.

Die Implementierung von **automatisierten Reaktionssystemen** durch KI-Algorithmen ermöglicht eine unmittelbare Identifizierung von Bedrohungen ohne menschliche Intervention, um zeitnah auf erkannte Gefahren zu reagieren.

Minimierung von Fehlalarmen durch die Integration von KI werden Fehlalarme reduziert werden, da das System lernt, normale von abweichenden Mustern zu unterscheiden.

Verbesserte **automatisierte Reaktionssysteme** gewährleisten eine sofortige Reaktion auf erkannte Bedrohungen, um Schäden zu begrenzen.

Kontinuierliche Anpassung der KI-Algorithmen passen sich fortlaufend neuen Angriffsmustern an, was die Genauigkeit der Bedrohungserkennung stetig verbessert.

Die Implementierung eines leistungsfähigen Monitoringsystems für KI-Anwendungen ist entscheidend für die Sicherstellung der Zuverlässigkeit und Effizienz von KI-Systemen in verschiedenen Branchen. Insbesondere das Monitoring von Cyberangriffen in Netzwerksystemen mittels KI-Technologie ist ein zentraler Bestandteil der modernen Sicherheitsarchitektur. Es spielt eine wesentliche Rolle dabei, Organisationen vor hochentwickelten Bedrohungen zu schützen und ihre Sicherheitsmaßnahmen zu stärken.

3.4.1. KI in Monitoring bei der Cyber-Überwachungen

Die zunehmende Digitalisierung und Vernetzung in der industriellen Produktion, auch bekannt als **Industrial Internet of Things (IIoT)** und **Industrie 4.0**, eröffnen neue Angriffsflächen für Cyber-Attacken. Isolierte Produktionsnetze werden geöffnet, was Wirtschaftsspionage und -sabotage begünstigt. Die Sicherheit von Produktionsnetzen (Production Operational Technology - Production-OT) wird durch die Zusammenlegung von Office-IT und Production-OT immer bedeutender, besonders in verschiedenartigen Produktionslandschaften.

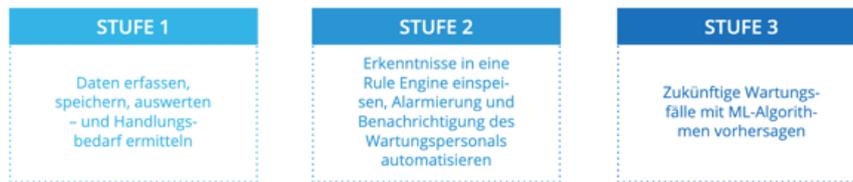


Abbildung 119: IOT & Maschinelles Lernen in 3 Stufen ML-Ansatz

Für eine schnelle Reaktion auf Sicherheitsvorfälle in industriellen Systemen wird das **Intrusion Detection and Prevention Systeme (IDPS)** herangezogen, um KI-Funktionalität, wie Anomalieerkennung und Schadsoftwaredatenströme zeitnah zu erkennen. Dabei unterstützt das Security Operation Center (SOC) die Verbesserung der Cybersicherheit in Unternehmen und stellt sehr hohe Qualitätsansprüche zur Verhinderung von Cyberangriffen, Erkennung von Schwachstellen und schnelle Reaktion auf interne/externe Bedrohungen. Dabei knüpfen weitere Systeme wie SIEM, SOAR, NDR, EDR und XDR mit KI-Funktionen, an und verbessern somit die Trefferergebnisse von Anomalien und Schadsoftware sowie ermöglichen eine schnellere betreiberspezifische Anpassung.

Erkennung von Anomalien in Datenströmen und Malware	Reaktion auf Anomalien in Datenströmen und auf erkannter Malware
SIEM	SOAR
<ul style="list-style-type: none"> • Erkennt Anomalien • Löst Warnung aus • Personal muss handeln • Manuelle Alarm-Triage 	<ul style="list-style-type: none"> • Nimmt Warnungen entgegen • Reagiert automatisiert • Trifft Entscheidungen, um Bedrohungen zu stoppen • Automatische Alarm-Triage

Abbildung 120: wesentliche Unterscheidungsmerkmale zwischen SIEM und SOAR

Digital Security Twin ist ein automatisiertes System für Risikobewertungen, das sehr hohe Qualitätsansprüche fordert und durch IT-Spezialisten mit KI Spezialwissen Weiterentwicklungen die Einsetzung von Alarm-Triage, Fehlalarme zu minimieren.

Herkömmliche **IDPS-Lösungen** filtern nur bekannte Angriffe heraus, hingegen **Generative Adversarial Networks (GAN)** Angriffe, die IDPS-Schwachstellen leicht erkennen, besitzen vielschichtige Abwehrsysteme und bessere Anpassungen von IDPS-Einstellungen sowie immer kontinuierliche Weiterentwicklungen von Sicherheitsmaßnahmen.¹³²

Auch Angreifer setzen auf KI und ML ein, um DDoS-Angriffe mit selbstlernenden Algorithmen und Verhaltensanalysen zu den Ziel-Systemen in Unternehmen oder Behörden durchzuführen. Es erfolgt eine ständige Angriffsmethodenstrategie von Angreifern oder Angegriffenen und wird mit herkömmlichen Gegenmassnahmen und Spezialwissen nicht mehr abgedeckt. Deshalb werden Massnahmen, wie der **Einsatz von Cloud-basierten Lösungen** mit sehr guten gefilterten, extrem hohen eingehenden Datenverkehr, die analysiert und blockiert werden können, bevor das IT-System im Unternehmen eingebunden wird.¹³³

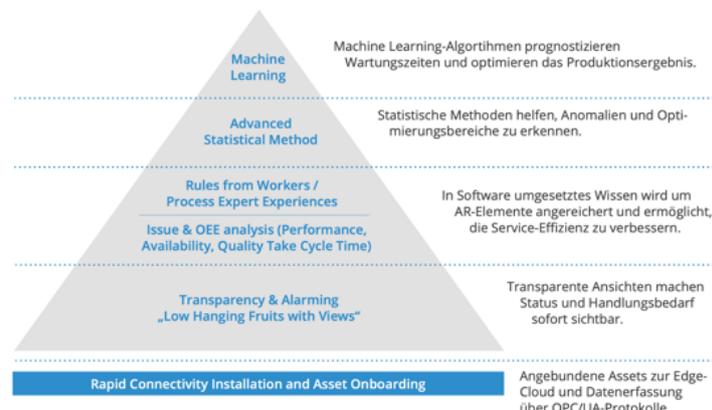


Abbildung 121: Rapid Connectivity Installation and Asset Onboarding

vgl.¹³² (Houdeau, 2023)

vgl.¹³³ (Sinner, 2020)

Rapid Connectivity Installation und Asset Onboarding in der Edge-Cloud mit OPC/UA-Protokolle schnellere Integration und effiziente Datenverbindung mit benutzerfreundliche Schnittstellen und minimale Ausfallzeiten, ermöglicht eine hervorragende Konnektivitätsgesamtleistung und digitale Transformationserleichterungen, **Transparenz und automatisierten Alarmierungsverbesserung** auf Sicherheitsereignissen mit schnelleren Reaktionsfähigkeit bei negativen Sicherheitsereignissen. Diese Fokussierung auf "**Low Hanging Fruits**" wird notwendige und schnellere Maßnahmen priorisieren und eine schnellere Reaktion auf die Sicherheitslagen erheblich verbessern.

Regeln basierend auf Erfahrungen und AR-Elemente mit der Integration von Regelwerken aus Mitarbeitererfahrung in Software mit AR-Elementen steigern die Service-Effizienz, indem präzise Anleitungen in Echtzeit bereitgestellt werden.

Fortgeschrittene statistische Methoden durch die Identifikation von Anomalien und Optimierungsbereichen durch präzise Datenanalysen, die eine proaktive Fehlerbehebung ermöglichen. Die Integration von KI verstärkt die Wirksamkeit.

Maschinelles Lernen in der Produktion mit Vorhersage von Wartungszeiten und die Optimierung des Produktionsergebnisses durch Machine Learning-Algorithmen verbessert die Effizienz, reduziert Ausfallzeiten und steigern die Qualität der hergestellten Produkte.¹³⁴

3.4.2. DDos

Schnelle Lokalisierung von DDoS-Angriffen durch den Einsatz von KI erfordert eine sorgfältige und systematische Vorgehensweise. Zunächst ist es wichtig, umfassende Daten über den normalen Netzwerkdatenverkehr zu sammeln, einschließlich Protokolldaten, IP-Adressen, Paketmetriken und Verbindungsinformationen. Diese Daten dienen als Grundlage für das Training von KI-Modellen.

Nach **Erstellung von Trainingsdaten** beginnt der eigentliche Schritt des Modelltrainings. Unterschiedliche KI-Algorithmen wie neuronale Netzwerke oder Entscheidungsbäume können dabei verwendet werden. Ziel ist es, das Modell auf das normale Verhalten des Netzwerks zu trainieren, um Muster und Merkmale von regulärem Datenverkehr zu erkennen.

Implementierung von KI-gesteuerten Überwachungstools ermöglicht eine Echtzeitanalyse des aktuellen Netzwerkdatenverkehrs. Diese Tools vergleichen kontinuierlich die aktuellen Aktivitäten mit dem vordefinierten Normalverhalten und nutzen KI für die Erkennung von Anomalien.

Zusätzlich werden **Schwellenwerte für bestimmte Metriken festgelegt**, wie beispielsweise die Anzahl der Anfragen pro Sekunde oder die Datenübertragungsrate. Bei Überschreitung dieser Schwellenwerte können automatisch Alarme ausgelöst werden, um auf potenzielle DDoS-Angriffe aufmerksam zu machen.

Kontinuierliche Verhaltensanalyse ermöglicht die Identifikation untypischer Aktivitäten von Benutzern oder Geräten. KI kann dabei Muster erkennen, die auf DDoS-Angriffe hindeuten, wie etwa repetitive Anfragen von denselben IP-Adressen.

Überwachung von IP-Adressen auf verdächtige Aktivitäten, kombiniert mit automatisierten Reaktion auf erkannte DDoS-Angriffe, stellt einen weiteren wichtigen Schritt dar. KI kann hier bekannte schädliche IP-Adressen identifizieren, blockieren, um das Netzwerk zu schützen. Abschließend wird eine **Feedbackschleife** implementiert, um das KI-Modell kontinuierlich zu verbessern. Durch die Aktualisierung basierend auf neuen Daten und sich entwickelnden Angriffsmustern bleibt die Sicherheitsinfrastruktur stets auf dem neuesten Stand. Diese ganzheitliche Vorgehensweise ermöglicht eine effektive und schnelle Lokalisierung von DDoS-Angriffen unter Einsatz von KI.¹³⁵

3.4.3. Ransomware und Viren

Die **Lokalisierung von Ransomware und Viren**, insbesondere solchen, die von künstlicher Intelligenz (KI) unterstützt werden, erfordert eine umfassende Sicherheitsstrategie. Ein mehrschichtiger Ansatz ist entscheidend, um die vielfältigen Angriffsvektoren abzudecken.

vgl.¹³⁴ (NIEWEG)

vgl.¹³⁵ (RCDev, 2023)

Verhaltensanalyse und Anomalieerkennung spielen eine zentrale Rolle. Sicherheitssysteme setzen auf fortschrittliche Analyseverfahren, um das normale Verhalten von Systemen und Netzwerken zu verstehen. Anomalien, die auf eine mögliche Infektion hinweisen, werden sorgfältig identifiziert. Hierbei kommt der Unterstützung durch künstliche Intelligenz besondere Bedeutung zu, da sie komplexe Muster und ungewöhnliche Verhaltensweisen schneller und präziser erkennen kann.

Signaturen und Heuristiken sind traditionelle Werkzeuge im Arsenal der Antivirensoftware. KI-gestützte Systeme gehen jedoch einen Schritt weiter, indem sie dynamisch neue Signaturen erstellen können, wodurch die Erkennung von Malware, insbesondere neuer Varianten, erschwert wird.

Die **Überwachung des Netzwerkverkehrs** ist ein weiterer entscheidender Aspekt. KI kann dazu beitragen, verdächtige Muster im Datenverkehr zu identifizieren, die auf eine mögliche Bedrohung hindeuten. Dies ermöglicht eine frühzeitige Reaktion und Eindämmung von potenziellen Angriffen.

Die **Verhaltensanalyse von Dateien**, insbesondere Erkennen von Aktivitäten wie dem heimlichen Verschlüsseln von Dateien durch Ransomware, ist ein Schlüsselfaktor. Hierbei greifen Sicherheitslösungen auf KI zurück, um verdächtige Aktivitäten präzise zu identifizieren.

Sandbox-Technologien bieten eine isolierte Umgebung, in der Dateien sicher ausgeführt werden können, um ihr Verhalten zu beobachten. KI unterstützt dabei, auch komplexe Bedrohungen in der Sandbox zu erkennen und zu analysieren.

Die **Anwendung von KI in der Mustererkennung**, im Code, im Verhalten von Anwendungen oder in der Netzwerkkommunikation, ermöglicht eine präzisere Identifizierung von Malware.

Schwachstellenmanagement ist von entscheidender Bedeutung, und KI kann bei der Identifizierung von Schwachstellen und der Priorisierung von Sicherheitsupdates eine unterstützende Rolle spielen.

Die **kontinuierliche Verbesserung der Sicherheitsmechanismen** ist unerlässlich, da sich Angriffstechniken ständig weiterentwickeln. Sicherheitslösungen selbst müssen daher KI-gestützt sein, um flexibel auf sich wandelnde Bedrohungen reagieren zu können.^{136 137}

Es ist wichtig zu betonen, dass die Bekämpfung von KI-gestützten Bedrohungen ein fortlaufender Prozess ist, der kontinuierliche Forschung und Anpassung erfordert, um die Sicherheitsmechanismen wirksam zu gestalten.¹³⁸

3.4.4. Phishing Angriffe

Künstliche Intelligenz (KI) spielt eine entscheidende Rolle in der **Abwehr von Phishing-Attacken** durch vielfältige Anwendungen:

Echtzeit-Erkennung von Phishing-Versuchen durch KI-Systeme überwachen den eingehenden Datenverkehr in Echtzeit, um Phishing-Versuche zu identifizieren. Sie erkennen verdächtige Aktivitäten in E-Mails und anderen Kommunikationskanälen.

Verhaltensanalyse durch die Überwachung von Nutzer- und Systemverhalten erkennt KI ungewöhnliche Aktivitäten, insbesondere solche, die auf Phishing-Angriffe hindeuten. Dies ist besonders nützlich, um Angriffe zu identifizieren, bei denen sich ein Angreifer als autorisierte Person ausgibt.

Filterung von Phishing-E-Mails mit KI-basierte Filter werden eingesetzt, um Phishing-E-Mails in Echtzeit zu erkennen. Diese werden automatisch in den Spam-Ordner verschoben, bevor sie die Posteingänge der Benutzer erreichen.

Erkennung von gefälschten Websites mit KI kann gefälschte Websites und betrügerische URLs erkennen, die von Phishern genutzt werden, um Nutzer zu täuschen.

Spracherkennung und Natural Language Processing (NLP) mit KI-Systeme analysieren Sprache und Text, um Phishing-Angriffe in verschiedenen Sprachen zu identifizieren.

vgl.¹³⁶ (SoSafe, 2023)

vgl.¹³⁷ (Acronis, 2023)

vgl.¹³⁸ (RCDev, 2023)

Historische Datenanalyse werden durch die Analyse historischer Daten von Phishing-Angriffen identifiziert KI Muster und Trends, um vor neuen Angriffen zu warnen.

Benutzer-Training werden durch den Einsatz von KI können Benutzer in der Erkennung von Phishing-Angriffen geschult und sensibilisiert werden, um menschliche Fehler zu minimieren.

Automatisierte Reaktion durch KI kann in Echtzeit auf Phishing-Angriffe reagieren, indem beispielsweise verdächtige E-Mails blockiert oder andere Maßnahmen ergriffen werden, um den Schaden zu begrenzen.

Es ist wichtig zu betonen, dass KI-Systeme nicht fehlerfrei sind und Phishing-Angriffe weiterhin eine ernsthafte Bedrohung darstellen. Eine effektive Abwehr erfordert oft die Kombination von KI und menschlicher Intelligenz. Dennoch stellen KI-Technologien eine bedeutende Ergänzung zu den Sicherheitsmaßnahmen von Unternehmen und Organisationen dar.¹³⁹

3.4.5. Zero-Day-Exploits und Schwachstellenerkennung

Die fortschrittliche Integration künstlicher Intelligenz (KI) in die Welt der Cyberkriminalität hat in den letzten Jahren zu einer alarmierenden Zunahme von **Zero-Day-Exploits und hochentwickelten Schwachstellenerkennungsmechanismen** geführt.

KI-Algorithmen haben die Fähigkeit erlangt, **Systeme in rekordverdächtiger Zeit zu analysieren**. Ihr Hauptziel besteht darin, ungesicherte Eintrittspunkte zu identifizieren und automatisierte Angriffe zu initiieren, die herkömmliche Sicherheitsmaßnahmen mit einer beunruhigenden Leichtigkeit umgehen können.

Besonders bedenklich ist die **Fähigkeit von KI-gesteuerten Angreifern**, bisher unbekannte Schwachstellen, die als Zero-Day-Exploits bezeichnet werden, eigenständig zu entdecken und auszunutzen. Diese Exploits sind besonders gefährlich, da für sie noch keine spezifischen Abwehrmechanismen existieren.

Genetische Algorithmen, die von den Prinzipien der natürlichen Selektion inspiriert sind, ermöglichen es der Malware, eine Vielzahl von Kandidatenlösungen zu durchlaufen. Dabei kommen genetische Operationen wie Mutation, Crossover und Selektion zum Einsatz. Dies ermöglicht eine Optimierung auf das Ziel hin, nämlich die Identifikation und Ausnutzung von Schwachstellen, bevor diese durch Sicherheitspatches oder Abwehrmechanismen geschlossen werden können.

Die **Anwendung von maschinellem Lernen** erlaubt es der KI-gesteuerten Malware, aus früheren Angriffen zu lernen und ihre Vorgehensweise kontinuierlich zu optimieren. Durch die Analyse von historischen Daten kann die Malware Muster erkennen, die auf potenzielle Schwachstellen hinweisen, und ihre Angriffsstrategien entsprechend anpassen. Diese adaptive Fähigkeit macht es äußerst schwer, sich gegen KI-gesteuerte Angriffe zu verteidigen, da sie sich dynamisch an veränderte Umgebungen anpassen kann.

Die **potenziellen Auswirkungen von Zero-Day-Exploits**, die von KI-gesteuerten Angreifern genutzt werden, sind erheblich. Da diese Angriffe auf bisher unbekanntem Schwachstellen basieren, können sie Sicherheitssysteme und -protokolle umgehen. Dies gefährdet sensible Daten, Netzwerke und kritische Infrastrukturen.

Der Zeitvorsprung, den KI-Angreifer durch die **Ausnutzung von Zero-Day-Exploits** erlangen, stellt eine ernsthafte Herausforderung für die Entwicklung und Implementierung effektiver Sicherheitsmaßnahmen dar.

Angesichts dieser Entwicklungen ist es von entscheidender Bedeutung, dass Unternehmen und Organisationen proaktiv handeln. Dies erfordert die Implementierung fortschrittlicher Sicherheitslösungen, die speziell auf KI-gesteuerte Bedrohungen abzielen. Eine kontinuierliche Überwachung, Anpassung und Weiterentwicklung von Sicherheitsstrategien ist erforderlich, um der sich ständig verändernden Bedrohungslandschaft effektiv zu begegnen und Zero-Day-Exploits rechtzeitig zu identifizieren und zu neutralisieren.¹⁴⁰

vgl.¹³⁹ (Huber, 2023)

vgl.¹⁴⁰ (RCDev, 2023)

3.4.6. RCDevs als eine Sicherheitsstrategie

RCDevs verfolgt eine umfassende Sicherheitsstrategie, die auf dem Zero-Trust-Ansatz basiert und dabei fortschrittliche Technologien wie **Multi-Faktor-Authentifizierung** (MFA) und leistungsstarke **Identity and Access Management** (IAM) Software integriert, um sich gegen eine Vielzahl von KI-gestützten Cyberangriffen zu verteidigen.

Im **Unterschied zu traditionellen Authentifizierungsmethoden setzt RCDevs auf einen innovativen MFA-Ansatz**, der die physische Welt einbezieht, z.B. durch die Verknüpfung der Authentifizierung mit der physischen Anwesenheit autorisierter Personen, was durch mobile App Tokens und Geolokalisierung ermöglicht wird. Diese Maßnahme schafft eine zusätzliche Sicherheitsebene und minimiert das Risiko böswilligen Zugriffs erheblich.

Die **MFA-Techniken von RCDevs** stellt für Angreifer mit KI-Technologien entgegen mit neuen Sicherheitsanforderungen, wie Verbindung von Wissen (Passwort), Besitz (mobile App Token) und Identität (Biometrie und Geolokalisierung) neue Sicherheitsanforderungen erfordert, um einen umfassenden und zuverlässigen Schutz vor fortschrittlichen Bedrohungen wie Phishing-Angriffen und KI-gestützter Malware. Entgegenzuwirken.

Die **OpenOTP Security Suite**, entwickelt von RCDev, reduziert das Risiko von unbefugtem Zugriff und Datendiebstahl und setzt auf die Implementierung von Multi-Faktor-Authentifizierungstechnologien wie **One-Time Passwords** (OTP), Mobile Push, **Fast Identity Identity Online2** (FIDO2), Sprachbiometrie und PKI sowie gelten nicht nur für Technologien mit Fernzugriff über VPN und WLAN, sondern auch für verschiedene Anmeldeszenarien wie Windows-Anmeldung und **Single Sign-On** (SSO).

Die **Überprüfung der Benutzeridentität** unterliegt mehreren Faktoren und setzt eine hohe Hürde für Angreifer, die gestohlene Anmeldeinformationen aus Phishing-Kampagnen nutzen möchten und dabei sind E-Mail-Clients und Webanwendungen mit erhöhen Schutz ausgestattet, um zu verdächtige Anmeldeversuche erkannt und zusätzliche Überprüfungsschritte erzwungen werden, was zu einer Echtzeitblockierung unbefugter Zugriffsversuche führt.

OpenOTP bekämpft **nicht direkt DDoS- und Brute-Force-Angriffe**, aber trägt indirekt dazu bei, dass Anmeldeversuche begrenzt sind und Zeitverzögerungen zwischen Fehlern eingeführt wurden. Dadurch wird das Risiko verringert, dass KI-gestützte Algorithmen systematisch Passwörter erraten oder entschlüsseln können und zusätzlich bietet OpenOTP verschiedene sichere Gegenoptionen gegen Brute-Force-Angriffe für den Second Factor.

Die **kostenlose OpenOTP Token App** legt grossen Wert auf maximale Sicherheit, da beim der Prozessbeginn durch einen QR-Code erkannt wird und für potenzielle Angreifer somit nutzlos ist, da durch biometrische Verifizierung und kryptographisches Schlüsselmanagement, Geolokalisierungserkennung, die bei dem betrügerische Verbindungsversuche, einschließlich Phishing-Angriffe ermittelt, werden, da alle Prozess auf den RCDevs Servern durchgeführt werden und so die Sicherheit weiter erhöhen lässt.

Die **OpenOTP Security Suite** senkt die Angriffsfläche durch die Verwendung eines "präsenzbasierten logischen Zugriffs". Das bedeutet, dass der Netzwerkzugriff gesperrt bleibt, es sei denn, ein Mitarbeiter hat sich ordnungsgemäß über das mobile Token angemeldet. Dadurch wird der Zutritt auf autorisierte Personen an autorisierten Standorten beschränkt.¹⁴¹

Investitionen in proaktive Sicherheitsmaßnahmen wie OpenOTP sind entscheidend im Kampf gegen KI-gestützte Cyber-Bedrohungen. Durch proaktives Handeln und die Implementierung notwendiger Maßnahmen können Einzelpersonen und Organisationen einen Schritt voraus sein, um die Integrität und Vertraulichkeit digitaler Vermögenswerte zu bewahren.

vgl.¹⁴¹ (RCDev, 2023)

3.5. KI/ML positive und negative Auswirkungen auf Monitoring

3.5.1. KI/ML positive Auswirkungen auf Monitoring

Folgende positive Auswirkungen auf Monitoring wurde nach Security Insider ermittelt:¹⁴²

Erkennung von Betrug und Anomalien der KI-Tools im Bereich der Cybersicherheit spielen eine entscheidende Rolle bei der Erkennung von komplexen Betrugsmustern und liefern durch den Einsatz von zusammengesetzten KI-Engines werden herausragende Ergebnisse geliefert. Dabei helfen fortschrittliche Analyse-Dashboards und diese bieten umfassende Details zu Betrugsfällen, insbesondere im Bereich der Anomalieerkennung.

E-Mail-Spam-Filter durch ML-basierte Spam-Filter sind in der Lage, Nachrichten mit verdächtigen Wörtern zu filtern und gefährliche E-Mails zu identifizieren und diese Filter schützen nicht nur vor schädlichen Inhalten, sondern reduzieren auch die Zeit, die für die Durchsicht unerwünschter Korrespondenz benötigt wird.

Die **Botnet-Erkennung** erzeugt von überwachten und unüberwachten maschinellen Lernalgorithmen eine frühzeitige Erkennung, um effektive Abwehr komplexer Bot-Angriffe durchzuführen und diese Algorithmen helfen, Verhaltensmuster von Nutzern zu analysieren und bisher unentdeckte Angriffe mit einer äußerst geringen False-Positive-Rate zu identifizieren. Dadurch wird die Präzision und Effizienz bei der Bekämpfung von Botnet-Bedrohungen erheblich verbessert.

Schwachstellen-Management bei der Verwaltung von Schwachstellen kann herausfordernd sein, aber KI-Systeme erleichtern diese Aufgabe erheblich, wie z.B. durch die Analyse von Benutzerverhalten, Endpunkte, Server und sogar Diskussionen im Dark Web können KI-Tools potenzielle Schwachstellen identifizieren und Angriffe vorhersagen.

Anti-Malware kann mit KI Unterstützung in Antivirensoftware zwischen guten und schlechten Dateien unterscheiden und ermöglicht die Identifizierung neuer Formen von Malware, selbst wenn sie zuvor noch nie gesehen wurden. Die Integration von KI-basierten Verfahren beschleunigt die Erkennung, wobei die Kombination mit traditionellen Methoden eine nahezu vollständige Erkennung aller Malware ermöglicht.

Verhinderung von Datenlecks werden durch KI-Erkennungen bestimmte Datentypen in Text- und Nicht-Text-Dokumenten untersucht und dabei können trainierbare Klassifikatoren darauf programmiert werden, verschiedene sensible Informationstypen in verschiedenen Formaten zu erkennen. Diese KI-Ansätze ermöglichen die Suche nach Daten in Bildern, Sprachaufzeichnungen oder Videos mithilfe geeigneter Erkennungsalgorithmen.

SIEM und SOAR mit ML-Tools können Sicherheitsinformations- und -ereignismanagement (SIEM) sowie Sicherheitsorchestrierung, -automatisierung und -reaktion (SOAR) integriert werden. Dadurch verbessert sich die Datenautomatisierung, die Sammlung von Informationen und die Erkennung verdächtiger Verhaltensmuster, während Reaktionen je nach Input automatisiert werden.

In verschiedenen Bereichen, darunter Netzwerkverkehrsanalyse, Intrusion Detection Systeme, Intrusion Prevention Systeme, Secure Access Service Edge sowie Benutzer- und Entitätsverhaltensanalysen, wird KI/ML erfolgreich eingesetzt. Moderne Sicherheitswerkzeuge ohne irgendeine Form von KI/ML sind heutzutage schwer vorstellbar.

vgl.¹⁴² (Andrey Shklyarov, Dmitry Vyrostkov)



Abbildung 122: Responsible AI Principles von infotech

3.5.2. KI/ML negative Auswirkungen auf Monitoring

Folgende negative Auswirkungen auf Monitoring wurden nach Security Insider ermittelt.¹⁴³

Sammeln von Daten durch den Einsatz von ML und Social Engineering können Cyberkriminelle Opferprofile besser erstellt werden, was zu beschleunigten Angriffen führen kann. Beispiele waren Botnet-Infektionen auf WordPress-Websites im Jahr 2018, die den Angreifern Zugang zu persönlichen Nutzerdaten verschafften.

Ransomware erlebt eine Renaissance und hat in einigen Fällen zu erheblichen Störungen geführt. Ein markantes Beispiel war die sechstägige Schließung der Colonial Pipeline und die Zahlung von 4,4 Millionen Dollar Lösegeld.

Spam, Phishing und Spear-Phishing können mithilfe von maschinellen Lernalgorithmen gezielt gefälschte Nachrichten erzeugen, die täuschend echt wirken und darauf abzielen, die Anmeldedaten von Nutzern zu stehlen. Eine Black-Hat-Konferenz zeigte, dass ein ML-Algorithmus in der Lage war, virale Tweets mit gefälschten Phishing-Links zu generieren, die viertel so effektiv waren als von Menschen erstellte Phishing-Nachrichten. Dies verdeutlicht das wachsende Risiko durch den Einsatz von KI zur Automatisierung und Optimierung von Cyberangriffen.

Fälschungen von Voice-Phishing nutzt die Deepfake-Audiotechnologie, die von ML generiert wird, um erfolgreiche Angriffe durchzuführen und die modernen Algorithmen können in wenigen Sekunden die Stimme einer Person analysieren und reproduzieren, was zu erfolgreichen Voice-Phishing-Angriffen führen kann.

Malware kann ML dazu verwendet werden, Malware zu verbergen, indem es das Verhalten von Knoten und Endpunkten imitiert. Moderne Algorithmen sind darauf trainiert, Daten schneller zu extrahieren als Menschen, was die Prävention von Malware erschwert.

Passwörter und CAPTCHAs können durch neuronale netzbasierte Software zunehmend leicht überwunden werden, indem sie menschliche Erkennungssysteme täuschen. ML ermöglicht es Cyberkriminellen, umfangreiche Passwort-Datensätze zu analysieren und effizientere Techniken zur Passwortvorhersage zu entwickeln, wie z.B. PassGAN, ein ML-basiertes Tool, das Passwörter präziser erraten kann als traditionelle Passwort-Cracking-Methoden. Dies unterstreicht die zunehmende Bedrohung durch den Einsatz von KI im Bereich der Cyberkriminalität.

Angriffe auf KI/ML selbst durch Missbrauch von Algorithmen, die in zentralen Sektoren wie Gesundheitswesen und Militär eingesetzt werden, könnten zu katastrophalen Folgen führen. Das Berryville Institute of Machine Learning hat eine Risikoanalyse von Angriffen auf ML-

vgl.¹⁴³ (Andrey Shklyarov, Dmitry Vyrostkov)

Algorithmen durchgeführt, was verdeutlicht, dass Sicherheitsingenieure lernen müssen, wie sie ML-Algorithmen in jeder Phase des Lebenszyklus sichern können.^{144 145}

Die Diskussion um den Einsatz von KI/ML in der Cybersicherheit sollte neben den offensichtlichen Vorteilen auch die Herausforderungen und Risiken berücksichtigen. Die effektive Nutzung dieser Technologien erfordert daher nicht nur ihre Integration in Sicherheitsmaßnahmen, sondern auch ein kontinuierliches Monitoring und die Anpassung an neue Bedrohungen.

3.5.3. KI Vergleiche Anwendungsleistungsmanagement vs. IT Operations Analytics

Application Performance Monitoring (APM) und **IT Operations Analytics (ITOA)** sind zwei Schlüsselbereiche im Bereich der Unternehmens-IT, die darauf abzielen, die Leistung, Verfügbarkeit und Effizienz von Anwendungen und IT-Infrastrukturen zu optimieren. Obwohl beide Disziplinen das gemeinsame Ziel der Verbesserung der IT-Performance teilen, gibt es wesentliche Unterschiede in ihren Ansätzen und Anwendungsbereichen.

APM konzentriert sich auf die Überwachung, Messung und Optimierung der Leistung von Anwendungen in Echtzeit und es bietet detaillierte Einblicke in den End-to-End-Transaktionspfad einer Anwendung, beginnend beim Endbenutzer bis hin zu den Backend-Servern, die die Kernziele von APM umfassen:

End User Equipment (EUE) bezeichnet eine Überwachung der Leistung aus der Perspektive der Benutzerinteraktion, um Engpässe oder Probleme aufzudecken, die sich auf die Benutzererfahrung auswirken könnten.

Transaktionsüberwachung verfolgt einzelne Transaktionen innerhalb einer Anwendung, um Engpässe zu identifizieren und die Antwortzeiten zu optimieren.

Infrastrukturüberwachung wird überwacht von Servern, Datenbanken und anderen Komponenten, um die Gesamtleistung der Anwendung zu gewährleisten.

Diagnose und Fehlerbehebung identifizieren Fehler, Engpässen oder andere Problemen, um schnell reagieren sowie die Anwendungsleistung optimieren zu können.

APM ist somit besonders relevant für Unternehmen, die eine Vielzahl von Anwendungen betreiben und sicherstellen möchten, dass diese Anwendungen den Erwartungen der Benutzer entsprechen.

ITOA hingegen fokussiert sich auf die Analyse großer Mengen von IT-Operationsdaten, um Muster, Anomalien und Trends zu identifizieren und anders als APM, das sich stark auf die Leistungsüberwachung von Anwendungen konzentriert, zielt ITOA darauf ab, umfassendere Einblicke in die Funktionsfähigkeit und Effizienz der gesamten IT-Infrastruktur zu liefern.

Big Data-Analyse erfolgt durch die Verarbeitung und Analyse großer Datenmengen, um verborgene Muster und Trends zu identifizieren.

Proaktive Fehlererkennung wird durch Früherkennung von Problemen und potenziellen Ausfällen, bevor diese sich auf die Anwendungsleistung auswirken.

Kapazitätsplanung wird benutzt für Prognosen von Ressourcenbedarf und Optimierung der Infrastruktur, um Engpässe zu vermeiden.

Automatisierte Reaktion durch die Implementierung von Automatisierung, um auf erkannte Probleme zu reagieren und die Systemintegrität aufrechtzuerhalten.

ITOA ist besonders relevant für Unternehmen mit komplexen IT-Infrastrukturen, da es eine ganzheitliche Sicht auf die gesamte Umgebung ermöglicht und Proaktivität bei der Fehlerbehebung und Ressourcenoptimierung fördert.

In der Praxis werden APM und ITOA oft kombiniert, um eine umfassende Überwachungs- und Analysestrategie zu schaffen, die sowohl die Anwendungsleistung als auch die Funktionsfähigkeit der gesamten IT-Infrastruktur adressiert. Unternehmen, die diese beiden An-

vgl.¹⁴⁴ (Andrey Shklyarov, Dmitry Vyrostkov)

vgl.¹⁴⁵ (Group, 2021)

sätze integrieren, sind besser gerüstet, um die Herausforderungen der modernen digitalen Landschaft zu bewältigen und eine optimale IT-Performance sicherzustellen.

Weitere Technologien und Anwendungen, laut Gartner, sind:¹⁴⁶

AI TRISM bezeichnet das Management von Vertrauen, Risiken und Sicherheit in Bezug auf KI.

CTEM bezieht sich auf die kontinuierliche Verwaltung von Bedrohungen und Risiken.

Intelligente Anwendungen verwenden KI-Technologien, um Benutzererfahrungen zu personalisieren und Probleme zu lösen.

Demokratisierte generative KI ermöglicht die Erstellung von kreativem Inhalt ohne umfangreiche technische Kenntnisse.

Technologieunterstützte vernetzte Belegschaft integriert Technologie in Arbeitsumgebungen zur Verbesserung der Zusammenarbeit.

Maschine Customers sind Organisationen, die KI-Systeme nutzen, um ihre Produkte oder Dienstleistungen zu verbessern.

3.6. KI-Tools

Einige KI-Tools werden beschrieben:¹⁴⁷

OpenAI Codex für Softwareprogrammierung.

Tabnine ist ein fortschrittliches Code-Vervollständigungstool.

Amazon Codewispherer ist eine automatische Codegenerierung.

BlackBox AI ist eine fortschrittliche Analyseplattform, um Daten tiefgründig zu analysieren.

Leonardo AI ist eine automatische Bildkorrektur und Optimierung.

PandaDoc ist eine umfassende Plattform für das Dokumentenmanagement mit Erstellung und Anpassung von Dokumenten.

Akkio implementiert KI und maschinellem Lernen in Unternehmen.

Adept AI implementiert KI-Algorithmen mit tiefgehender Datenanalyse, um menschliche Fähigkeiten zu erweitern und zu nutzen.

ChatGPT Sprachassistent ist eine natürliche Sprachverarbeitung.

Galileo AI Automatisierte Erstellung von UI/UX-Designs.

ChatGPT mit umfangreichen Trainingsdaten Erstellung von Texten zu einer Vielzahl von Themen und weitere ...¹⁴⁸

3.7. KI-Verordnung der EU

Die Europäische Union verabschiedete am 13. März 2024 mit der Mehrheit den ersten AI-Act der Welt, der in 01. August 2024 in Kraft tritt.¹⁴⁹ **Rechtsrahmen für den Einsatz** vertrauenswürdiger Künstlicher Intelligenz (KI) zielen darauf ab, die ethische Nutzung von KI-Systemen zu gewährleisten und gleichzeitig die grundlegenden Menschen- und Verbraucherrechte zu schützen. Der EU AI-Act ist eine Verordnung zur Regelung von KI-Systemen und sieht Übergangsfristen von 6 bis 36 Monaten vor, je nach Risikoklasse. Der zusammengefasste AI-Acts EU findet sich im Excel File AI-Act-EU.xlsx.



Abbildung 123: Laufe des AI-Acts von 2024 bis 2026

vgl.¹⁴⁶ (McCartney, 2023)

vgl.¹⁴⁷ (Universe, 2024)

vgl.¹⁴⁸ (Universe, 2024)

vgl.¹⁴⁹ (KOMMISSION, 2024)

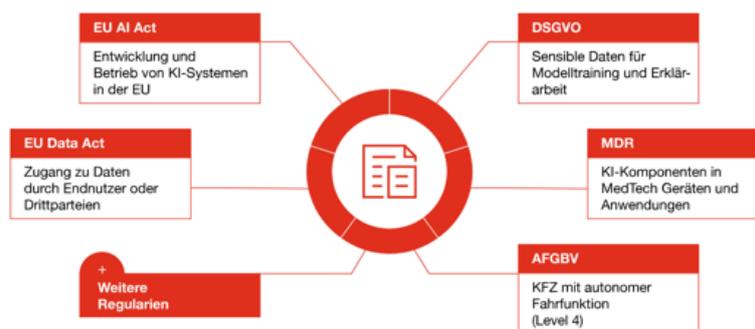


Abbildung 124: KI-Regulierungen und ihre Umsetzungen

Die KI-Systeme werden nach Risiken reguliert, einige sind verboten, andere müssen bestimmte Anforderungen erfüllen, wobei die Hauptlast die Nutzer und Anbieter von KI-Systemen mit hohem Risiko tragen. **General Purpose AI** (GPAI) unterliegen strengerer Regeln, besonders bei Transparenz und Urheberrecht. KI-Systeme, die mit Menschen interagieren, müssen diese über ihre Nutzung informieren. Neben dem AI-Act der EU gibt es weitere Vorgaben, wie z.B. die Datenschutz-Grundverordnung (DSGVO) oder den Data-Act EU sowie branchenspezifische Vorschriften, wie z.B. die Medizinprodukteverordnung (MDR) oder die deutsche autonome-Fahrzeuge-Genehmigungs- und Betriebs-Verordnung (AFGBV).¹⁵⁰

Der **Geltungsbereich des KI-Gesetzes** bezieht sich auf Unternehmen, die auf dem EU Markt KI-Systeme anbieten, deren KI-Systeme von EU Nutzern verwendet oder deren KI-generierte Inhalte innerhalb der EU als auch Staaten außerhalb der EU genutzt werden. Die Nichteinhaltung dieser Vorschriften kann zu erhebliche Geldstrafen von bis zu 35 Millionen Euro oder 7% des Jahresumsatzes und zu Haftungsrisiken führen.

Das **Gesetz umfasst technische und administrative Anforderungen** an KI-Systeme und berücksichtigt dabei vier Risikostufen (gering, normal, hoch und inakzeptabel) eingeteilt. Dabei sollen KI-Anwendungen entsprechend ihrem Gefahrenpotenzial reguliert und überwacht werden, sodass diese ethisch und rechtskonform sicher eingesetzt werden.

Der **AI-Act hat grundsätzlich Auswirkungen auf sämtliche KI-Systeme**, die unter die Definition von KI fallen sind maschinenbasierte Systeme, die in erheblichem Maße autonom arbeiten, Vorhersagen treffen, Empfehlungen aussprechen, Texte generieren oder Entscheidungen treffen, die sich auf physische oder virtuelle Umgebungen auswirken. Einige KI-Systeme bleiben unberührt, da ihr Output außerhalb der EU genutzt wird, sie nicht auf dem EU Markt agieren, ausschließlich für private Zwecke, Forschung oder militärische Anwendungen genutzt werden oder das System erfüllt nicht die Definition von Künstlicher Intelligenz gemäß dem AI-Act EU. Dennoch ist es wichtig, die Gesetzgebung kontinuierlich anzupassen, da zukünftige Änderungen Auswirkungen auf die Sicherheit der Gesellschaft haben könnten.

Unakzeptables Risiko ist grundsätzlich nicht akzeptabel, außer in wenigen Ausnahmefällen, die für die strafrechtliche Verfolgung erforderlich sind, wie z.B. subliminale Techniken, die potenziell Schaden verursachen können oder zwei KI -Systeme, die miteinander agieren und außer Kontrolle geraten.

Hohes Risiko kann akzeptabel sein, sofern die gesetzlichen Anforderungen eingehalten und eine Compliance-Überprüfung durchgeführt wird, wie z.B. KI-Systeme im Bildungsbereich, bei der Beschäftigung, Einwanderung oder im Rechtswesen.

Limitiertes Risiko kann zugelassen werden, solange definierte Informations- und Transparenzstandards eingehalten werden, wie z.B. Chatbots und Deep Fakes.

Geringes Risiko hingegen ist ohne Einschränkungen zulässig, obwohl die Einhaltung eines Verhaltenskodex empfohlen wird, wie z.B. Spamfilter, Predictive Maintenance und dynamische Preisgestaltung.¹⁵¹

vgl.¹⁵⁰ (Reese)

vgl.¹⁵¹ (GmbH)

Hochrisiko-KI-Systeme nach Annex I (vormals Annex II)		
Abschnitt A – Hochrisiko-KI-Systeme für die, zur Vermeidung von doppelter regulatorischer Belastung, Flexibilität bei der Umsetzung der Compliance eingeräumt wird:		
<ul style="list-style-type: none"> • Maschinen • Sicherheit von Spielzeugen • Sportboote und Wassermotormäder • Aufzüge und Sicherheitsbauteile für Aufzüge 	<ul style="list-style-type: none"> • Geräte und Schutzsysteme in explosionsgefährdeten Bereichen • Funkanlagen • Druckgeräte • Seilbahnen 	<ul style="list-style-type: none"> • Persönliche Schutzausrüstungen • Geräte zur Verbrennung gasförmiger Brennstoffe • Medizinprodukte • In-vitro-Diagnostika
Abschnitt B – Hochrisiko-KI-Systeme, die sektoral reguliert, jedoch von den meisten Anforderungen ausgenommen sind:		
<ul style="list-style-type: none"> • (Sicherheit in der) Zivilluftfahrt • Zwei-, drei- und vierrädrige Fahrzeuge • Land- und forstwirtschaftliche Fahrzeuge 	<ul style="list-style-type: none"> • Schiffsausrüstung • Interoperabilität des Eisenbahnsystems • Kraftfahrzeuge und Kraftfahrzeuganhänger sowie Systeme, Bauteile und 	<ul style="list-style-type: none"> • selbstständige technische Einheiten für diese Fahrzeuge
Hochrisiko-KI-Systeme nach Annex III		
<ul style="list-style-type: none"> • Biometrische Systeme • Kritische Infrastruktur • Bildungswesen • Personalwesen 	<ul style="list-style-type: none"> • Essenzielle private und öffentliche Dienstleistungen inklusive Finanzen und Versicherungen • Strafverfolgung 	<ul style="list-style-type: none"> • Migration, Asyl und Grenzkontrollen • Rechtspflege und demokratische Prozesse

Abbildung 125: Hochrisiko-KI-Systeme nach Annex I und III

Die AI-Acts basieren auf der Erkenntnis, dass es bisher keine klare Definition von "künstlicher Intelligenz" gibt und gesetzliche Regulierungssysteme und Anforderungen erforderlich sind, um die rasch voranschreitende Technologie in der Schnittstelle zwischen Mensch und Technik, verbotene Praktiken, zu regulieren. Dazu gehören die Entwicklung und Nutzung von KI-Systemen, die Personen manipulieren, physischen, psychischen Schaden zufügen können, aber auch bei KI-Systeme wird untersagt, dass die Schwäche oder Schutzbedürftigkeit bestimmter Gruppen ausnutzen und Personen schädigen können sowie Social-Scoring-Systeme zur Bewertung der Vertrauenswürdigkeit von Personen sind verboten. Massnahmen sind notwendig, um auch der Cyberkriminalität wirksam entgegenzuwirken. Nach Art. 84 KI-VO-EU der Verordnung zielt darauf ab, einheitliche Standards festzulegen und den Einsatz von KI-Systemen verantwortungsvoll zu lenken, wobei sie sich auf eine Reihe von Bausteinen des Bundesamts für Sicherheit in der Informationstechnik (BSI) stützt, mit Ausnahme unter¹⁵².

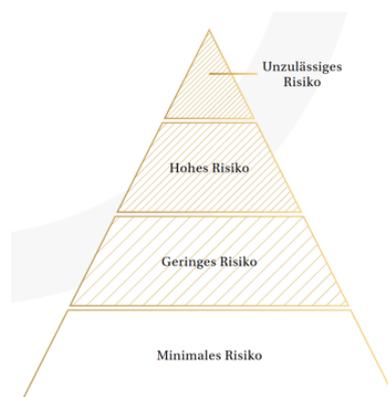


Abbildung 126: KI-Risiko Pyramide

Unakzeptables Risiko und Hochrisiko von KI-Systeme stellen den Schwerpunkt des AI-Act EU dar, da sie ein erhöhtes Risiko für die Gesundheit, Sicherheit und Grundrechte der EU-Bürger:innen darstellen. Unternehmen nutzen bereits heute möglicherweise Anwendungen, die in Zukunft als Hochrisiko-KI eingestuft werden könnten. Die Einstufung einer KI als Hochrisiko hängt davon ab, ob diese potenziell eine Gefahr für Einzelpersonen darstellt.

vgl.¹⁵² (BSI)

Die Verordnung unterscheidet auch KI-Systeme mit **geringem und minimalem Risiko**. **KI-Systeme mit geringem Risiko** sind für die Interaktion mit Menschen gedacht und unterliegen nicht den strengen Regeln für verbotene oder Hochrisiko-KI. Unternehmen, die solche Systeme entwickeln oder nutzen, müssen jedoch Transparenzanforderungen erfüllen. **KI mit minimalem Risiko** stellen keine expliziten Risiken dar und erfordern keine speziellen Verpflichtungen gemäß der Verordnung, obwohl die Förderung von Verhaltenskodizes für Anbieter dieser Systeme vorgesehen ist.

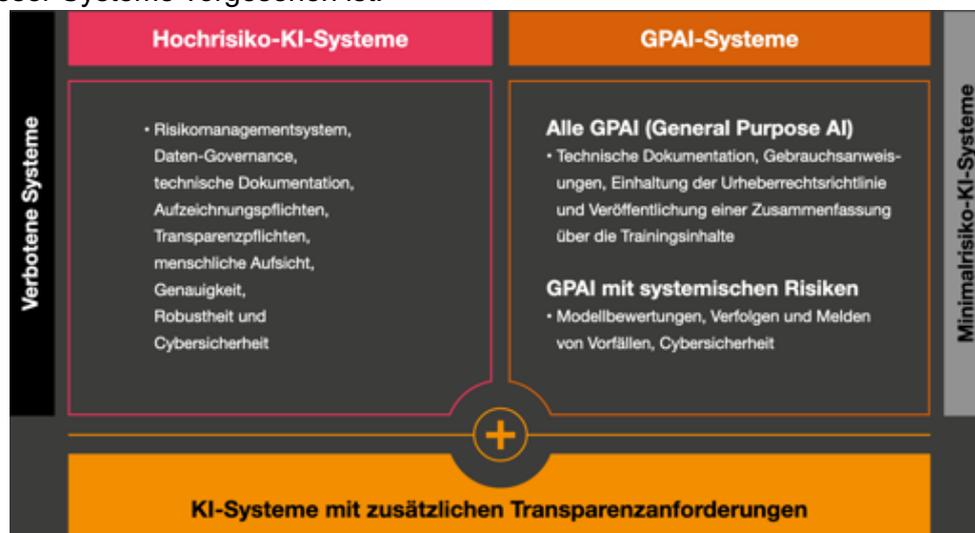


Abbildung 127: Kategorien von KI-Systemen nach dem EU AI-Act

Der AI-Act EU klassifiziert KI-Systeme basierend auf ihrem Risikopotenzial in verschiedene Kategorien: **uneingeschränkt zulässige**, unter bestimmten Auflagen **erlaubte** und **verbotene Systeme**. KI-Systeme mit **untragbaren Risiken**, wie etwa solche zur unterschwellig- en Manipulation von Menschen oder zur biometrischen Echtzeitüberwachung im öffentlichen Raum, sind strikt untersagt. Der Schwerpunkt der Verordnung liegt jedoch auf **Hochrisiko-KI-Systemen**, die zwar zugelassen sind, jedoch strenge Anforderungen in Bezug auf Dokumentation, Überwachung und Qualitätssicherung erfüllen müssen. Diese Hochrisiko-Systeme decken eine Vielzahl branchenspezifischer Anwendungsfälle ab und unterliegen besonderer regulatorischer Kontrolle. Zuletzt wurden KI-Systeme mit allgemeinem Verwendungszweck (GPAI) in die Verordnung aufgenommen. Anbieter von GPAI-Systemen müssen technische Dokumentationen, Gebrauchsanweisungen und Urheberrechtsrichtlinien bereitstellen sowie Informationen zum Training veröffentlichen. GPAI-Systeme mit systemischen Risiken müssen zusätzliche Evaluierungen durchführen, Risiken eindämmen, Vorfälle verfolgen und die Cybersicherheit gewährleisten. KI-Systeme, die direkte oder indirekte Auswirkungen auf menschliche Nutzer haben, unterliegen erhöhten Transparenzpflichten, z.B. bei biometrischer Kategorisierung oder Medienmanipulation. KI-Systeme ohne spezifische Risiken sind ohne zusätzliche Anforderungen zulässig, aber die EU-Kommission kann die Liste der regulierten Systeme aktualisieren und weitere Anwendungsfälle hinzufügen.¹⁵³

Anlehnend von Herrn Schürmann Rosenthal Dreyer beurteilt, die Gem. Art. 28 KI-VO-EU, die **Pflichten für Anbieter von Hochrisiko-KI** und **Pflichten für Nutzer von Hochrisiko-KI**:¹⁵⁴

Anbieter von Hochrisiko-KI-Systemen unterliegen einer Reihe von Pflichten gemäß der KI-Verordnung und diese müssen sicherstellen, dass ihr System alle Anforderungen gemäß den Artikeln 8–15 der Verordnung erfüllt. Sollten diese Anforderungen nicht erfüllt werden, sind diese verpflichtet, Korrekturmaßnahmen gemäß Artikel 16 lit. g der Verordnung durchzuführen. Darüber hinaus müssen diese gegenüber der zuständigen nationalen Behörde jederzeit nachweisen können, dass diese die erforderlichen Maßnahmen ergriffen haben (Artikel 16 lit. j). Weiterhin müssen Anbieter von Hochrisiko-KI-Systemen ein Qualitätsmanagementsystem gemäß Artikel 17 der Verordnung einrichten (Artikel 16 lit. b). Diese

vgl.¹⁵³ (Dreyer)
vgl.¹⁵⁴ (Dreyer)

sind verpflichtet, Protokolle gemäß Artikel 12 der Verordnung aufzubewahren, sofern dies der Kontrolle des Anbieters unterliegt (Artikel 16 lit. d). Zur Gewährleistung der Konformität müssen diese ein Konformitätsbewertungsverfahren gemäß Artikel 19 durchführen und das KI-System gemäß Artikel 49 mit einer CE-Kennzeichnung versehen (Artikel 16 lit. e, i). Weiteren müssen die Anbieter das System gemäß Artikel 51 der Verordnung in der EU-Datenbank registrieren (Artikel 16 lit. f) und bei Nichtkonformität des Systems mit nationalen Behörden und notifizierten Stellen zusammenarbeiten (Artikel 16 lit. h). Diese Pflichten sind entscheidend, um die Sicherheit, Transparenz und Rechenschaftspflicht bei der Entwicklung und Nutzung von Hochrisiko-KI-Systemen zu gewährleisten.

Nutzer von Hochrisiko-KI-Systemen sind ebenfalls verpflichtet, bestimmte Maßnahmen gemäß der KI-Verordnung zu ergreifen. Zunächst müssen diese sicherstellen, dass die Eingabedaten den vorgesehenen Zweck des KI-Systems entsprechen (Artikel 29 Absatz 3) und diese verantwortlich für die Überwachung des Betriebs des KI-Systems gemäß den Angaben in der Gebrauchsanweisung (Artikel 29 Absatz 4) sind. Es obliegt den Nutzern, Protokolle gemäß Artikel 12 der Verordnung aufzubewahren, sofern dies ihrer Kontrolle unterliegt (Artikel 29 Absatz 5) und diese müssen die bereitgestellten Informationen gemäß Artikel 13 nutzen, um gegebenenfalls eine Datenschutz-Folgenabschätzung gemäß Artikel 35 der Datenschutz-Grundverordnung (DSGVO) durchzuführen. Diese Pflichten dienen dazu sicherzustellen, dass Nutzer von Hochrisiko-KI-Systemen die ordnungsgemäße Verwendung und den sicheren Betrieb dieser Systeme gewährleisten, besonders im Hinblick auf den Datenschutz und die Einhaltung der vorgesehenen Zwecke. Gesetzliche Anpassungen für die KI-Verordnung werden durch die EU auf ihre Online-Plattform mitgeteilt.¹⁵⁵

Laut PrivacyXperts¹⁵⁶ gelten Pflichten für Hochrisiko-KI ab 01.August 2024:

- Stand der Technik beachten
- Risikomanagementsystem / Qualitätsmanagementsystem (Art. 9, 17 KI-VO) muss vorhanden sein
- Resilienz / Cyber-Security (Art. 15 KI-VO)
- Testverfahren / regelmäßige Tests (Art. 16 KI-VO) müssen stattfinden
- EU-Konformitätserklärung und CE-Kennzeichnung (Art. 16 KI-VO) müssen vorliegen
- Zusammenarbeit mit den zuständigen Behörden (Art. 21 KI-VO)
- Korrekturmaßnahmen / Informationspflichten müssen umgesetzt werden (Art. 20 KI-VO)
- technische Dokumentation / Aufbewahrungspflichten (Art. 11, 18 KI-VO)
- Transparenz (Art. 13 KI-VO)
- Protokollierung von Funktionsmerkmalen (Art. 12 KI-VO)
- Beobachtung nach Markteinführung (Art. 71 KI-VO)
- Meldung von „schwerwiegenden Vorfällen“ (Art. 73 KI-VO)
- Durchführung einer Grundrechte-Folgenabschätzung (Art. 27 KI-VO)
- Entwicklung mit Trainingsdaten mit einer bestimmten Qualität (Daten-Governance, Art. 10 KI-VO)
- Registrierung (Art. 49 KI-VO)
- menschliche Aufsicht (Art. 14 KI-VO)
- Maßnahmen zur Barrierefreiheit (Art. 16 KI-VO)

weitere relevante Gesetze aus der KI-VO-EU sind:¹⁵⁷

- Risikomanagementsysteme (Art. 9 KI-VO-E) 21
- Daten und Data-Governance (Art. 10 KI-VO-E) 21
- Dokumentationspflichten (Art. 11 KI-VO-E) 22
- Whitepaper - Umsetzung der KI-VO in Unternehmen²
- Aufzeichnungspflicht (Art. 12 KI-VO-E) 23
- Transparenz- und Instruktionspflichten 23 (Art. 13 KI-VO-E)
- menschliche Aufsicht (Art. 14 KI-VO-E) 24

¹⁵⁵ (KOMMISSION, 2024)

¹⁵⁶ (PrivacyXperts)

¹⁵⁷ (Dreyer)

- Genauigkeit, Robustheit und Cybersicherheit 24 (Art. 15 KI-VO-E)
- Qualitätsmanagementsystem (Art. 17 KI-VO-E) 25
- Konformitätsbewertungsverfahren 25 (Art. 19 KI-VO-E)
- weitere Verpflichtungen für Hochrisiko-KI 27
- Transparenzpflichten für bestimmte KI-Systeme 27 (Art. 52 KI-VO-E)
- freiwillige Verhaltenskodizes für Systeme ohne 27 hohes Risiko (Art. 69 KI-VO-E)
- Folgen von Verstößen (Art. 71 KI-VO-E)

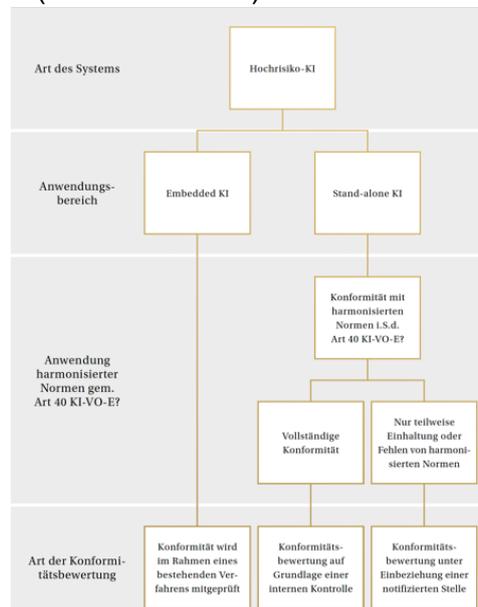


Abbildung 128: gesetzliche KI-Verordnung der EU

Die **KI-Verordnung** (KI-VO) 2024 legt wichtige Regelungen für die Nutzung von Künstlicher Intelligenz (KI) fest und setzt einen zentralen Aspekt für das Risikomanagement, das Anbieter dazu verpflichtet, kontinuierlich Risiken zu identifizieren, analysieren und entsprechende Maßnahmen zu ergreifen. Die Verwendung von Daten für die KI-Entwicklung unterliegt strengen Anforderungen in der Relevanz, der Repräsentativität und der Qualität, um Diskriminierung zu vermeiden.

Weitere zusammenfassende Bestimmungen betreffen die technische Dokumentation, automatische Protokollierung, Transparenz- und Instruktionspflichten für KI-Systeme. Dabei wird menschliche Aufsicht während des gesamten KI-Einsatzes gefordert, ebenso wie die Gewährleistung von Genauigkeit, Robustheit und Cybersicherheit sowie ein Qualitätsmanagementsystem und Konformitätsbewertungsverfahren sind ebenfalls vorgeschrieben.

Für **Hochrisiko-KI** gibt es zusätzliche Verpflichtungen wie die Aufbewahrung von Protokollen, die Durchführung von Korrekturmaßnahmen und die Zusammenarbeit mit nationalen Behörden sowie die Transparenzpflichten für KI-Systeme mit geringem Risiko sowie freiwillige Verhaltenskodizes für solche Systeme werden ebenfalls angesprochen. Bei den Verstößen gegen die Verordnung können mit empfindlichen Sanktionen, einschließlich hoher Bußgelder, geahndet werden.¹⁵⁸

Nationale KI-Strategien unterscheiden sich inhaltlich, aufgrund der unterschiedlichen rechtlichen und gesellschaftlichen Regelungen, wie der **OECD**, dem **AI-Watch des JRC**, dem **Future of Life Institute**, **AlgorithmWatch** und anderen. Der bevorstehende EU-Rechtsakt wird als innovativ betrachtet, da er einen umfassenden regulatorischen Rahmen für KI vorschlägt und andere Länder könnten die EU als Vorbild betrachten oder weniger geneigt sein, strikte Regulierungen weltweit für KI zu erlassen. Eine lebhaft diskutierte Möglichkeit der Gesetzgebung im Bereich KI finden in vielen Ländern statt und mit dem Ziel, trotz unterschiedlichen gesetzlichen Bestimmungen, wie Datenschutzgesetze in den einzelnen Staaten, eine weltweite KI-Verordnung anzustreben.¹⁵⁹

vgl.¹⁵⁸ (Dreyer)
vgl.¹⁵⁹ (COMMISSION, 2023)

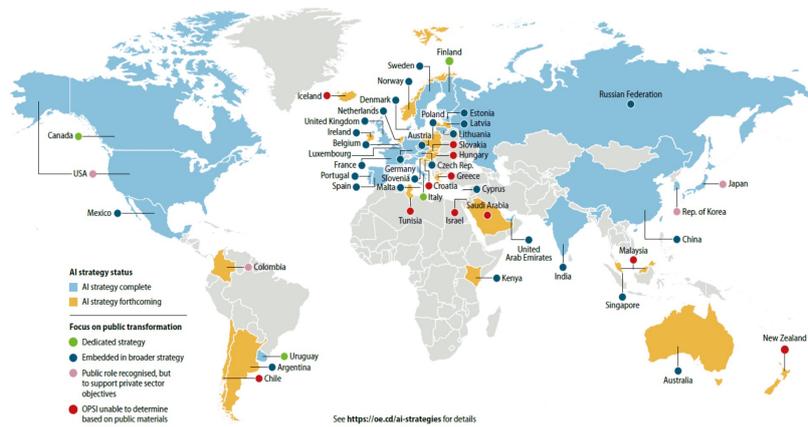


Abbildung 129: internationale Landschaft der KI-Initiativen

3.8. Risikoanalyse-Bewertung und Vorbeugung von KI



Abbildung 130: Risk Mitigation Essential

Unternehmen sollten Risikomanagement von KI von Anfang an berücksichtigen. Dies ermöglicht es Entwicklern, KI-Modelle gemäß den Unternehmenswerten und dem Risikoappetit zu erstellen, indem Werkzeuge wie Modellinterpretierbarkeit, Bias-Erkennung und Leistungsüberwachung integriert werden. Standards, Tests und Kontrollen sollten in den gesamten Lebenszyklus des Analysemodells eingebettet sein, von der Entwicklung bis zur Nutzung.



Abbildung 131: Lebenszyklus des Analysemodells

Typischerweise werden Risikokontrollen in der Analyse erst nach Abschluss der Entwicklung angewendet und dies kann zu Verzögerungen führen, wenn Probleme auftreten, die einen weiteren Entwicklungszyklus erfordern. Eine Lösung besteht darin, Risikoidentifikation und -bewertung sowie Kontrollanforderungen direkt in die Entwicklungs- und Beschaffungszyklen einzubetten, was die Vorimplementierungsprüfungen beschleunigt und die Verzögerungen reduziert. Organisationen müssen auch frühzeitig mit Kontroll- und Geschäftssteams sowie Anbietern zusammenarbeiten, um potenzielle Risiken zu verstehen und zu mindern. Durch die Integration von Kontrollen in die Entwicklungs- und Bereitstellungsabläufe können diese einen sichereren und agileren Ansatz zur Analyse verfolgen, wie z.B. kann ein solches Risiko sein, dass Voreingenommenheit in Daten und Analysemethoden bestehen, die zu ungerichteten Entscheidungen führen können. Führende Unternehmen integrieren daher verschiedene Kontrollen in ihre Analyse-Entwicklungsprozesse ein.

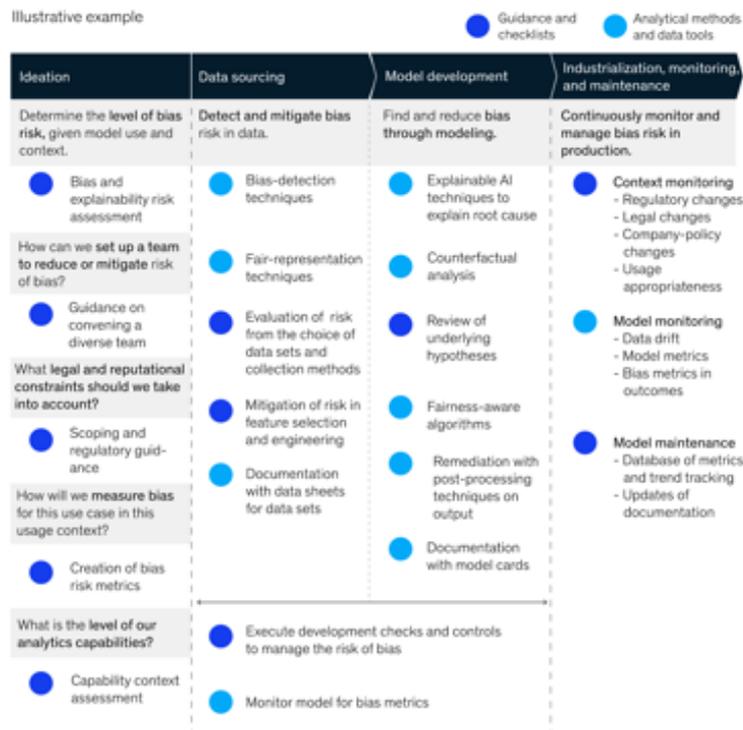


Abbildung 132: Risikoentwicklung

Ideenfindung im Geschäftsanwendungsfall werden zu Beginn und im regulatorischer und reputativer Kontext verstanden und beim frühzeitigen Erkennen der Risiken helfen passende Anforderungen an Daten und Methoden, die korrekt und sinnvoll festzulegen sind.

Datenbeschaffung mit Datensätzen sind oft Voreingenommenheiten enthalten, sollten ein Voreingenommenheitstests durchgeführt werden.

Modellentwicklung wirkt sich auf Transparenz und Interpretierbarkeit auf des Voreingenommenheitsrisikos und deshalb wählen Unternehmen geeignete Methoden und Techniken zur Erklärbarkeit aus.

Überwachung und Wartung legen Unternehmen Überwachungsanforderungen von Leistung fest, in Abhängigkeit vom Anwendungsfall und Aktualisierungshäufigkeit des Modells sowie die Unterstützung von automatisierte Technologieplattformen bei Überwachungstests-durchführungen.¹⁶⁰

Risiko und Risikomanagement für KI werden in 6-Schritte-Prozess unter Verwendung von algorithmischer Tools erstellt; zusätzliche Risikoquellen herausgefiltert; KI-Governance-Risiken - MLOPs/DLOPs für Versionierung, Validierung, Leistung und Überwachung von KI-Modellen durch verschiedene Tools und Governance-Praktiken; Benutzerkompetenz-Risiko für Schulungen und Coaching; Führungsrisiko durch Einstellung und Trägheit gegenüber Data Science und KI sowie Risikofehlentscheidungen durch fehlende Identifizierung und Analyse und Minderung.

vgl.¹⁶⁰ (Juan Aristi Baquero, 2020)

Step 1: Risikoklassifizierung

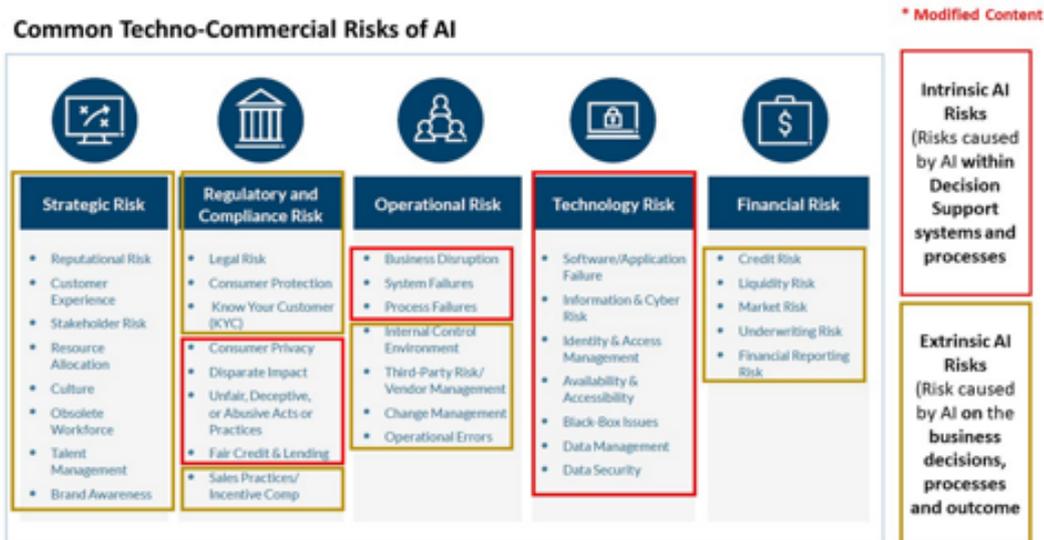


Abbildung 133: Risikoklassifizierung

Schritt 1: KI-Risiken verursacht interne geschäftliche Entscheidungsunterstützungssysteme und externe geschäftliche Entscheidungen zu klassifizieren, um ein entsprechende Risikominderungsmechanismen zu entwickeln. Ein solches Framework sollte technische und kommerzielle Aspekte berücksichtigen, um beide Risikoarten entsprechend zu kontrollieren.

Step 2: Risk Management Framework

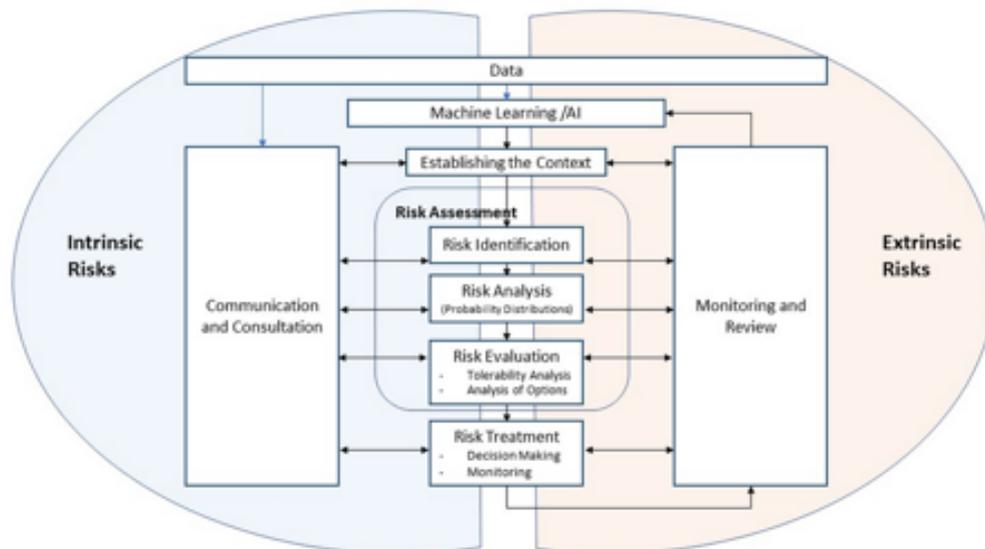


Abbildung 134: Risk Management Framework

Schritt 2: Risikomanagementrahmen für die Unternehmen und Behörden werden interne und externe Risiken erstellt für das KI Umfeld. Ein Forschungspapier von Žigienė¹⁶¹ schlägt einen 4-stufigen Risikomanagementrahmen nach **ISO 31000** vor:

- **Risikoerkennung:** Wahrscheinlichkeit negativer Ereignisse A basierend auf statistischen/ML- und KI-Methoden festlegen.
- **Risikoanalyse:** Erwartete negative Folgen quantifizieren.
- **Risikobewertung:** Priorisierung der Risiken nach ihrer Auswirkung.
- **Risikomanagement:** Entwicklung von Maßnahmen zur Kontrolle, Reduzierung und Minderung von Risiken.

vgl.¹⁶¹ (Žigienė, 2019)

Step 3: Bewertung der Maßnahmen zur Verwaltung intrinsischer Risiken



Abbildung 135: Risk Management Framework

Das Risikomanagement-Framework von Protiviti¹⁶² ist sehr umfassend und deckt Risiken in den Bereichen Daten, Modellierung, Testen, Betrieb und Governance ab und automatisiert ein 2-stufiges Überwachungssystem in drei Intelligenzebenen:

1. KI implementierte Modelle werden in Live-Geschäftsprozess eingesetzt und überwachen diese.
2. KI-Risikomodell verarbeitet Prozessdatenströme und vergleicht Ausgaben mit Normalwerten ab, um Differenzen zu bewerten und Risiken zu berechnen.
3. Dashboard zeigt Ereignisse von Risiken für verschiedene KI-Modelle an und ermöglicht deren Steuerung.

KI-Risikomanagementlösungen beinhalten statistische Tools, Analysen, Szenarien und ML-Modelle sowie Risikoeinschätzungen, um richtige Entscheidungen zu ermöglichen.

2-Level Algorithmic Risk Monitoring and Management

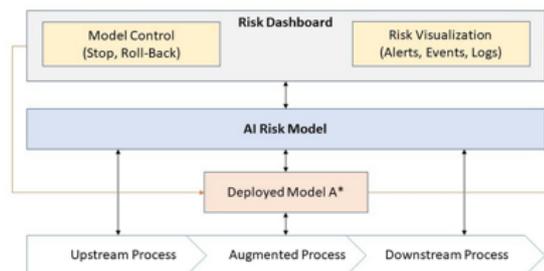


Abbildung 136: 2 Level Algorithmic Risk Monitor und Management

Schritt 4: Bewertung der Dimensionen und Maßnahmen zur Verwaltung extrinsischer Risiken

Žigienė¹⁶³ beschreibt KI-basiertes Risikomanagementmodell für Unternehmen und Behörden für externe Risiken hinsichtlich Konkurrenzdruck, Gruppenzwang, Angst, Anreize mit den automatisierten KI-System-Prozessen und dabei fließen Eingabedaten aus vorherigen Prozessvorhersagen in nachfolgende Prozesse ein, die letztendlich implementierte KI-Modells effektiv managt, negative Ergebnisse und Geschäftsentscheidungsrisiken erstellen, Prozesse und Ergebnisse vorhersagen. Unternehmen sollten frühzeitig Risikomodelle erarbeiten, um ein vollständiges Profilrisikomodelle von verschiedenen Risikoarten zu lokalisieren und die wichtigsten Risikotreiber zu definieren.

vgl.¹⁶² (Protiviti, 2022)

vgl.¹⁶³ (Žigienė, 2019)

ML/AI based Extrinsic Risk Management

* Modified Content

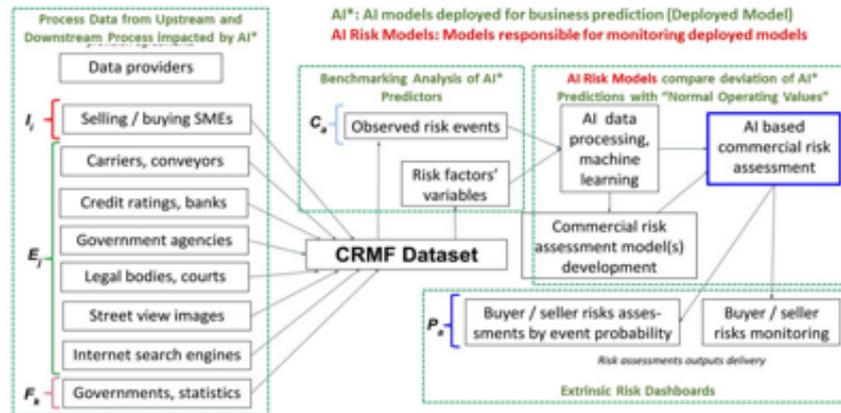


Abbildung 137: ML/AI basiert extrinsisches Risikomanagement

Schritt 5: Risikobewertung und Auswirkungsabschätzung

Algorithmic Risk Scoring and Impact Assessment

* Modified Content

Impact Level	Description	Definition	Score % Range
Level 1	Level I decisions will often lead to impacts that are reversible and brief.	Little or no impact	0% – 25%
Level 2	Level II decisions will often lead to impacts that can be difficult to reverse, and are ongoing	Moderate impact	26% - 50%
Level 3	Level III decisions will often lead to impacts that can be difficult to reverse, and are ongoing	High impact	51% - 75%
Level 4	Level IV decisions will often lead to impacts that are irreversible, and are perpetual.	Very high impact	76% - 100%

Consolidated Risk Area	No. of evaluation Criteria	Max. Score
1 – Process/Product	14	20
2 – System	12	15
3 – Algorithm	8	25
4 – Decision	10	15
5 – Business Impact	16	25
6 – Data	10	20
Raw Impact Score	70	120

Abbildung 138: Algorithmische Risikobewertung und Auswirkungsabschätzung

Interne Risiken, wie Störungen, werden durch technische Lösungen von Daten- und KI-Experten verwaltet, hingegen externe und restliche Risiken fordern kommerzielle Strategie.

Step 6: Risk Governance

Nach dem Schritt 5 werden Governance-Maßnahmen für jedes Risikolevel erstellt und die technischen Maßnahmen erkennt und umsetzt.¹⁶⁴

Risk Governance for Intrinsic and Extrinsic AI Impact

* Modified Content

	Level 1	Level 2	Level 3	Level 4
Decision Review	None	At least one of Qualified domain expert or a data and AI advisory		At least two of Qualified domain experts AND a data and AI advisory
Human in the Loop (HITL)	Decisions may be rendered without human involvement		<ul style="list-style-type: none"> Decisions cannot be made without human intervention points Final decision must be made by a human 	
Explanation Requirement	Meaningful explanation required for decision process e.g. FAQs	Meaningful explanation required for any decision that resulted in the denial of a benefit or service		
Training	None	Documentation on the design and functionality of the system	<ul style="list-style-type: none"> Documentation on the design and functionality of the system. Mandatory training of decision professionals 	<ul style="list-style-type: none"> Documentation on the design and functionality of the system. Re-occurring training of decision professionals Proof of training completion
Contingency Planning	None		Mandatory contingency plans and/or backup systems should the Automated Decision System be unavailable	
Approval for the system to operate	None	Functional Manager or equivalent	Business Unit Head or equivalent	Executive Board

Abbildung 139: Risk Governance

vgl.¹⁶⁴ (Gupta)

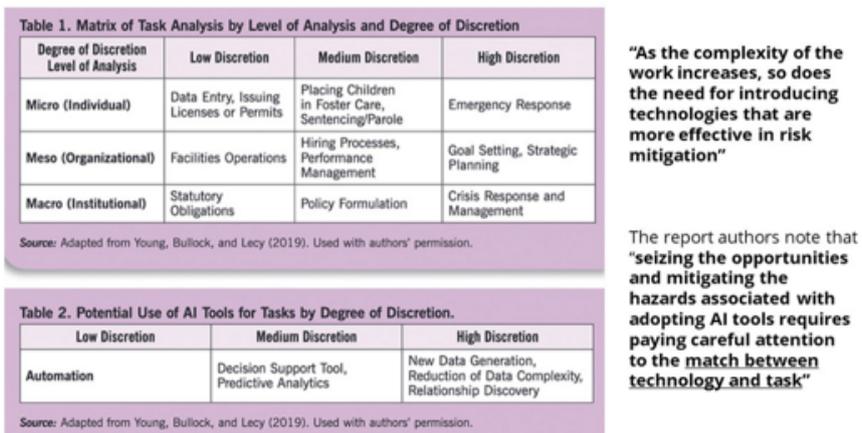


Figure 140: Künstliche Entscheidungsfreiheit als Grundlage für risikobasierte Governance

Die wirklichen Risiken entstehen aus der Interaktion der Modelle mit den realen Prozessen, die aus "Risikotoleranz" etabliert werden, um eine effektive geschäftliche Governance sicherzustellen.¹⁶⁵

3.8.1. Risikoanalyse-Erstellung

Der folgende Abschnitt präsentiert eine Risikoanalyse, die im Detail in Anhang I ausgeführt wird.

Die subjektive Betrachtung bezieht sich auf SOC- CMM for CERTs¹⁶⁶ und der AI-Act EU¹⁶⁷ zur Regulierung von Künstlicher Intelligenz in der EU.

Der Vergleich zwischen **Computer Emergency Response Teams (CERTs)**¹⁶⁸ und dem **AI-Act (KI-Verordnung)** der EU zeigt, dass beide unterschiedliche Aufgaben und Ziele haben und in bestimmten Bereichen zusammenhängen können. Während CERTs sich mit **Cyber-sicherheitsvorfällen** und dem Schutz von Informationssystemen befassen, regelt der AI-Act den **Einsatz von Künstlicher Intelligenz (KI)** und konzentriert sich auf die Sicherheit und den verantwortungsvollen Umgang mit KI-Technologien.

Der AI-Act EU zielt darauf ab, ein einheitliches Regelwerk für die KI in der EU zu schaffen, das die Sicherheit und die Grundrechte der Bürger schützt. KI-Anwendungen werden in verschiedene Risikokategorien eingeteilt: niedrige, mittlere, hohe und inakzeptable (sehr hohe) Systeme und dabei unterliegen die hohem und sehr hohen Risiken strengen Anforderungen, während niedrig-riskante Anwendungen weniger reguliert sind. Hersteller von KI-Systemen müssen umfangreiche Dokumentationen, Risikobewertungen und Sicherheitsprüfungen vorlegen, um die Konformität mit den Vorschriften sicherzustellen. Ein zentraler Aspekt des AI-Acts EU ist sind ethische Überlegungen, Datenschutz und den Schutz der Grundrechte der Bürger:innen. Die Durchsetzung erfolgt durch nationale Behörden in den EU-Mitgliedstaaten. Im Gegensatz dazu konzentriert sich **NIST**¹⁶⁹ auf die Entwicklung von Standards und Leitlinien zur Verbesserung der KI-Technologie, insbesondere in Bezug auf deren Sicherheit und Vertrauenswürdigkeit. Die NIST-Richtlinien befassen sich stark mit technischen Aspekten der KI, einschließlich der Entwicklung von Metriken und Bewertungsrahmen zur Messung von KI-Systemen. NIST verfolgt einen flexiblen Ansatz, der es Unternehmen ermöglicht, die besten Praktiken und Standards auszuwählen, die für ihre spezifischen Anwendungen geeignet sind. Die Organisation arbeitet eng mit der Industrie, Forschungseinrichtungen und anderen Regierungsbehörden zusammen, um eine breite Palette von Perspektiven und Expertise einzubeziehen und betont die Förderung von Innovation und Wettbewerbsfähigkeit in der KI-Branche.

vgl.¹⁶⁵ (Gupta, 2022)

vgl. ¹⁶⁶ ff. (SOC-CMM, 2024)

¹⁶⁷ (Union, 2024)

vgl.¹⁶⁸ (BSI)

¹⁶⁹ (NIST)

Trotz ihrer Unterschiede gibt es auch Ähnlichkeiten, zielen beide Ansätze darauf ab, die Sicherheit und Vertrauenswürdigkeit von KI-Systemen zu erhöhen und berücksichtigen ethische Aspekte, wenn auch in unterschiedlichem Umfang. Insgesamt sind der AI-Act der EU und die NIST-Richtlinien komplementäre Ansätze, die Unternehmen, die international tätig sind, dazu anregen, sowohl rechtliche Rahmenbedingungen als auch beste Praktiken zu berücksichtigen, um Compliance und Innovation zu gewährleisten.¹⁷⁰

Eine weitere wichtige Richtlinie ist die **NIS2-Richtlinie**¹⁷¹, die ein ergänzendes Regelwerk der AI-Act EU ist und auf die unterschiedlichen Aspekte der Digitalisierung abzielt. Die KI-Verordnung fokussiert sich auf den sicheren und verantwortungsvollen Einsatz von KI-Systemen, während NIS2 den Schutz kritischer Infrastrukturen und Cybersicherheitsstandards stärkt. Beide Verordnungen fördern die Sicherheit, Transparenz und Verantwortlichkeit im Umgang mit neuen Technologien und tragen dazu bei, die digitale Souveränität und Sicherheit in der EU zu gewährleisten.

Bewusst wurde von der herkömmlichen Risikomatrix abgewichen, ohne vorherige Gefährdungseinschätzung nach BSI, um die wichtigsten Zusammenhänge aller Komponenten anschaulicher darzustellen. Diese Gesamtanalyse umfasst die SOC-Bereiche Business, People, Process, Technology und Service. Im Mittelpunkt stehen die Risikoerkennung, -analyse, -bewertung und -bewältigung von Security Information and Event Management (SIEM)-Systemen, die auf Künstlicher Intelligenz (KI) basieren, innerhalb eines SOC-Systems¹⁷² auf der Grundlage von AI-Act EU¹⁷³ erstellt, welche Unternehmen und Behörden als Checkliste verwenden können.

In diesem Kontext werden spezifische Fragestellungen im Hinblick auf die relevanten Bestimmungen von Bereiche identifiziert und ausgearbeitet, insbesondere hinsichtlich der Risikoeinstufung, Gewichtigkeit, Wahrscheinlichkeit und Auswirkungen, Beschreibung und entsprechenden Gegenmaßnahmen.

Ziel ist es, eine grafische Darstellung der Ergebnisse auf Basis des SOC-CMM-Schemas¹⁷⁴ zu erstellen. Diese Darstellungen wurde durch Python-Analysen Tool Jupyter Notebook¹⁷⁵ und Spyder¹⁷⁶ sowie den Einsatz von ChatGPT¹⁷⁷ (Formulierungen) unterstützend erstellt und wissenschaftlich fundiert abgeglichen (siehe auch Quellenangaben in Excel-File). Die Analyse integriert Erkenntnisse aus verschiedenen Quellen, insbesondere aus den Abschnitten 3 und 4 dieser Masterarbeit eingebunden. Der Reifegrad (Maturity) der Technologien in allen fünf Bereichen sowie die organisatorischen, technischen und fachlichen Fähigkeiten (Capability) werden separat im Bereich Technology und Service anhand zentraler Aspekte untersucht. In den unterschiedlichen Bereichen wird der Maturity- und Capability-Score klar sichtbar, wobei die unterschiedlichen Bewertungen auf der Identifizierung relevanter Merkmale basieren. Bedarfsgemäss können zusätzliche relevante Faktoren in die Analyse einbezogen werden. Die subjektive Auswertung wurde in **SOC-AI.xlsx** unter den Tab «Next Steps» beschrieben (Quellenangaben in der Datei angegeben).

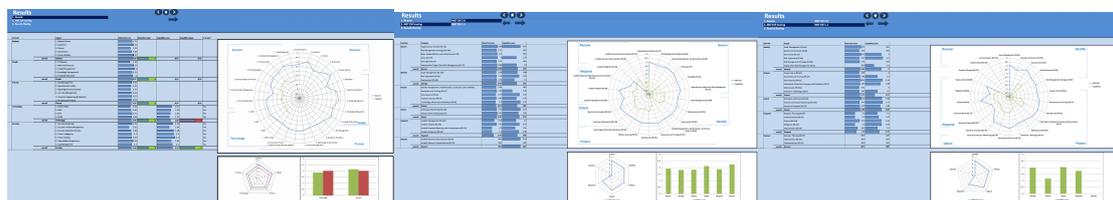


Abbildung 141. KI-Analyse

vgl.¹⁷⁰ (Kruger, 2024)

¹⁷¹ (Union, 2022)

vgl.¹⁷² ff. (SOC-CMM, 2024)

¹⁷³ (Union, 2024)

vgl.¹⁷⁴ ff. (SOC-CMM, 2024)

¹⁷⁵ ff. (Jupyter)

¹⁷⁶ ff. (Spyder)

¹⁷⁷ (ChatGPT)

Spezial zum Bereich „**Process 5. KI-Life-Cycle-Process**“ wurde die Betrachtung der KI-gesteuerte Lebenszyklus für KI-Systeme durchgeführt, welche alle Phasen umfassen von der Anforderungsanalyse bis hin zur kontinuierlichen Überwachung und Optimierung des Systems. Anlehnend wurden Fragestellungen von AIGA¹⁷⁸ herangezogen, um eine Risikobewertung vom KI-Life-Cycle-Process zu erstellen. Jede Phase ist entscheidend, um die Sicherheit der IT-Infrastruktur zu gewährleisten und Bedrohungen rechtzeitig zu erkennen. Weitere Betrachtungen können noch herangezogen werden. Durch eine gründliche Risikobewertung in jeder Phase können Fehlalarme reduziert und die Sicherheit verbessert werden. Diese Ausarbeitung wurde durch Python-Analysen Tool Jupyter Notebook¹⁷⁹ und Spyder¹⁸⁰ sowie den Einsatz von ChatGPT¹⁸¹ (Formulierungen) unterstützend erstellt und wissenschaftlich fundiert abgeglichen (siehe auch Quellenangaben in Excel-File), anlehnend zum AI-Act der AIGA¹⁸².

Separat wird eine KI-Betrachtungen gemäß dem AI-Act der EU¹⁸³ in den Bereichen Verordnung 1 - 180, Artikel 1 -136 und den Anhängen I - XIII durchgeführt, analysiert und bewertet, um eine Grafik zu erstellen. Die subjektive Auswertung wurde in **AI-Act-EU.xlsx** unter den Tab «Next Steps» beschrieben (Quellenangaben in der Datei angegeben). Diese Darstellungen wurden durch Python-Analysen Tool Jupyter Notebook¹⁸⁴ und Spyder¹⁸⁵ erstellt, anlehnend zum AI-Act der EU.¹⁸⁶

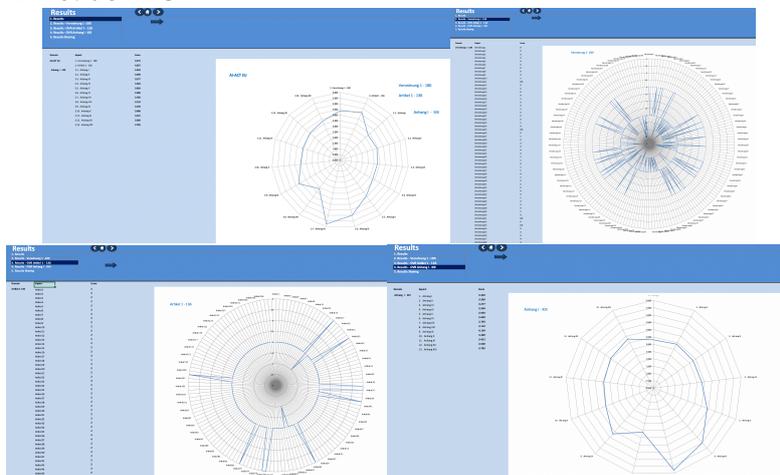


Abbildung 142: AI-Act-Analyse

Der AI-Act der EU und die NIST-Richtlinien sind ergänzende Ansätze zur Regulierung von KI. Während der AI-Act einen rechtlichen Rahmen mit strengen Anforderungen bietet, zielt NIST darauf ab, Standards und Praktiken zu fördern, die Innovation und Entwicklung zu unterstützen. Unternehmen, die international tätig sind, sollten beide Rahmenbedingungen berücksichtigen, um Compliance und beste Praktiken zu gewährleisten.

Hinweis: Die **Messung von KI-Risiken** ist subjektiv, da oft keine klaren Methoden existieren, die durch die Integration von Drittanbieter-Software oder -Daten verstärkt werden, und somit keine allgemein akzeptierten Methoden zur Bewertung von Risiken und Vertrauenswürdigkeit geben. Hierbei müssen Messungen den spezifischen Kontext und unterschiedliche Auswirkungen auf betroffene Gruppen berücksichtigen und angepasst werden. Unternehmen müssen ihre eigene **Risikotoleranz** definieren, abhängig von rechtlichen Anforderungen und dem spezifischen Anwendungsfall, und diese Toleranzbereiche sollten sich zeitlich ändern, da sich KI-Systeme und Richtlinien weiterentwickeln.

vgl.¹⁷⁸ ff. (AIGA)

¹⁷⁹ ff. (Jupyter)

¹⁸⁰ ff. (Spyder)

¹⁸¹ (ChatGPT)

¹⁸² ff. (Spyder)

vgl.¹⁸³ (Union, 2024)

¹⁸⁴ ff. (Jupyter)

¹⁸⁵ ff. (Spyder)

vgl.¹⁸⁶ (Union, 2024)

Bei der **Risikopriorisierung** können nicht alle KI-Risiken vollständig eliminiert werden, daher sollten Unternehmen die Ressourcen gezielt auf die größten wichtigsten Risiken verteilt werden, um Systeme mit sehr hohen und hohen Risiken bzw. mit Daten sensiblen Informationen arbeiten oder direkt Menschen betreffen, sollten priorisiert behandelt werden als geringere Prioritäten für weniger riskante Systeme.

Integration in Unternehmensstrategien von KI-Risiken sollten nicht isoliert betrachtet werden, sondern in bestehende Unternehmensrisikomanagementprozesse tiefgründig analysiert und im Zusammenhang mit anderen Risiken wie Cybersecurity und Datenschutz betrachtet werden.¹⁸⁷

3.9. Erstellung von gesetzlichen Anforderungen nach BSI-Schema

Die folgenden Bausteine werden annähernd gemäß dem Schema des BSI IT-Grundschutzkompendiums erstellt und beziehen sich auf einige KI-Eigenschaften und die Risikoanalysen vom KI-Lifecycle-Process. Diese Elemente werden anlehnd vom Abschnitt 4.8.1.2. und 4.8.1.3. (Anhang I) umfassend in Anhang II beschrieben. Bewertung erfolgte mit MUSS, SOLLTE, DARF, DARF NICHT (KANN wurde nicht betrachtet.)

Die ausführliche subjektive Beschreibung der Bausteine - KI-Eigenschaften erfolgt im Anhang II.

	 TECHNICAL DESIGN CHARACTERISTICS	 SOCIO-TECHNICAL CHARACTERISTICS	 GUIDING PRINCIPLES CONTRIBUTING TO TRUSTWORTHINESS
AI RMF Taxonomy	<ul style="list-style-type: none"> • Accuracy • Reliability • Robustness • Resilience or ML Security 	<ul style="list-style-type: none"> • Explainability • Interpretability • Privacy • Safety • Managing Bias 	<ul style="list-style-type: none"> • Fairness • Accountability • Transparency
OECD AI Recommendation	<ul style="list-style-type: none"> • Robustness • Security 	<ul style="list-style-type: none"> • Safety • Explainability 	<ul style="list-style-type: none"> • Traceability to human values • Transparency and responsible disclosure • Accountability
EU AI Act	<ul style="list-style-type: none"> • Technical Robustness 	<ul style="list-style-type: none"> • Safety • Privacy • Non-discrimination 	<ul style="list-style-type: none"> • Human agency and oversight • Data governance • Transparency • Diversity and fairness • Environmental and societal well-being • Accountability
EO 13960	<ul style="list-style-type: none"> • Purposeful and performance-driven • Accurate, reliable, and effective • Secure and resilient 	<ul style="list-style-type: none"> • Safe • Understandable by subject matter experts, users, and others, as appropriate 	<ul style="list-style-type: none"> • Lawful and respectful of our Nation's values • Responsible and traceable • Regularly monitored • Transparent • Accountable

Abbildung 143: KI-Eigenschaften, die auf Richtlinien dokumente abgebildet werden

Die ausführliche subjektive Beschreibung der Bausteine - KI-Lifecycle-Process erfolgt im Anhang II.

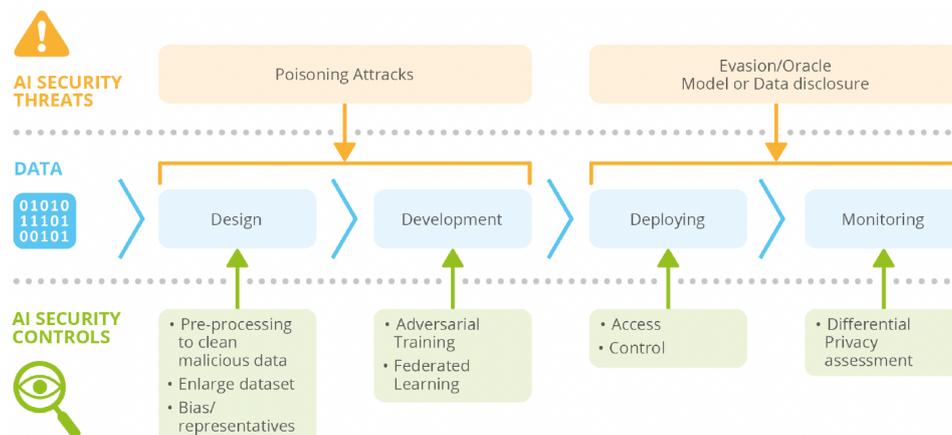


Abbildung 144: Beziehung zwischen KI-Bedrohungen und Sicherheitskontrollen

vgl.¹⁸⁷ (NIST, 2023)

3.10. KI Trends in der Informatik

Die rasante Entwicklung von Informationstechnologien hat bereits heute weitreichende Auswirkungen auf verschiedenste Branchen mit selbstlernende Systeme der Künstlichen Intelligenz (KI) werden komplexe Probleme eigenständig gelöst und dies ist nur ein Beispiel für den Einfluss der fortschreitenden Digitalisierung.

Künstliche Intelligenz (KI) und Maschinelles Lernen (ML) revolutionieren die Art und Weise, wie Systeme lernen, Entscheidungen treffen und Muster erkennen, wie z.B. personalisierten Dienstleistungen bis hin zu fortschrittlichen Datenanalysen verändert diese Technologie verschiedene Branchen, wo selbstlernende Systeme sind in der Lage, eigenständig komplexe Probleme zu lösen und tragen somit maßgeblich zur Weiterentwicklung verschiedener Sektoren bei.

Quantum Computing versprechen eine immense Steigerung der Rechenleistung, sodass komplexeste Berechnungen und Analysen ermöglichen und potenzielle transformative Auswirkungen auf Forschung, Kryptographie und Problemlösungen haben. Dies hat das Potenzial, grundlegende Veränderungen in der informatischen Landschaft herbeizuführen.

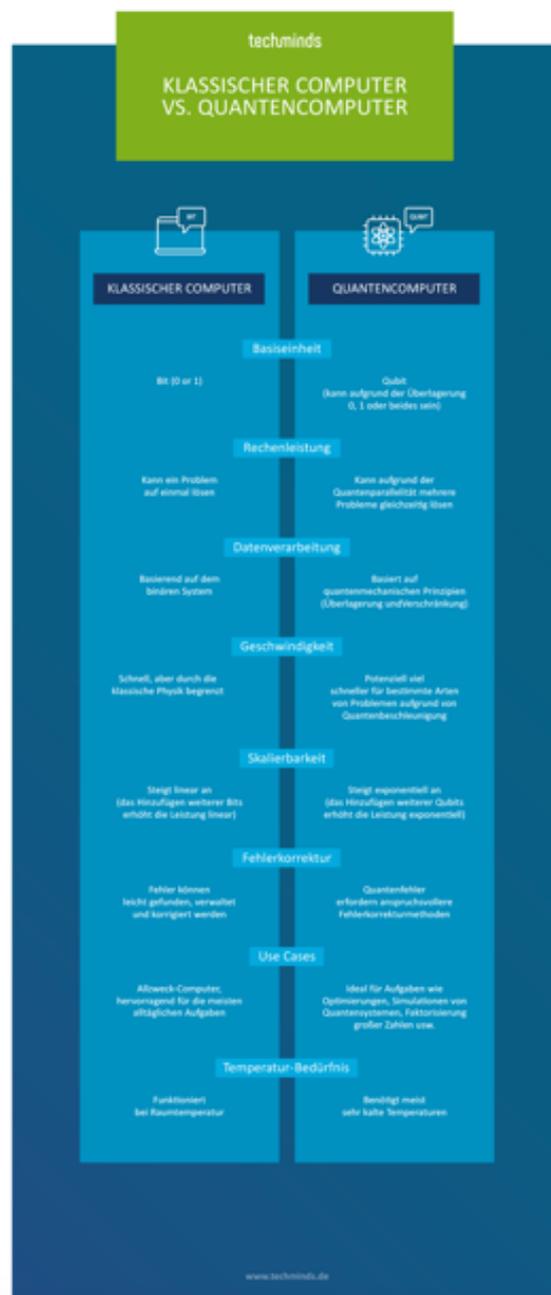


Abbildung 145: Vergleich klassische Computer vs. Quantenrechner

Augmented Reality (AR) und Virtual Reality (VR) bieten immersivere Erlebnisse in Bereichen, wie z.B. im Gaming, in der Bildung und im Training von Daten und Informationen und bieten somit den Technologien neue Möglichkeiten für virtuelle Interaktionen und Simulationen zu eröffnen.

Internet der Dinge (IoT) ist eine Vernetzung von Geräten im Internet der Dinge und ermöglicht einen reibungslosen Datenaustausch und die Automatisierung von Abläufen. Dabei beeinflusst dieser Trend alle Bereiche, wie z.B. im Smart Homes, Gesundheitswesen und Industrie.

5G und darüber hinaus ist eine Einführung von 5G ermöglicht schnellere und zuverlässigere drahtlose Verbindungen und beschleunigt somit den Mobilfunk, treibt die Entwicklung von Technologien im autonomen Fahren an, erweiterte Konnektivität, Virtual Reality, vernetzte Maschinen und bewegt eine Schlüsseltechnologie für zukünftige Innovationen voran.

Autonome Fahrzeuge mit integrierter KI, ML und 5G gewährleisten die sichereren Transportmittel von A nach B.

Blockchain-Technologie bietet dezentrale und sichere Datenspeicherung und hat Auswirkungen auf Finanztransaktionen, Lieferkettenmanagement, Urheberrechtsschutz, Authentifizierung und digitaler Identitätsverifizierung mit zunehmenden Anwendungen.

Cybersecurity bleibt eine zunehmende digitale Bedrohung und deshalb ist es wichtig, durch neue Fortschritte in der Abwehr von Bedrohungen und Datenschutz neue vernetzte Systeme zu schützen. Durch den Anstieg von Datenlecks steigt auch die Nachfrage nach innovativer IT-Sicherheit, die einen effektiven Schutz vor immer raffinierteren Angriffen bietet.

Human Augmentation ist eine Erweiterung körperlicher und kognitiver Fähigkeiten durch Technologie hat das Potenzial, die menschliche Erfahrung auf ein neues Level zu heben.

Automatisierung und Robotik bieten eine verbesserte Effizienz und Produktivität in verschiedenen Sektoren, vor allem beim autonomen Fahrzeugen bis hin zu automatisierten Fertigungsprozessen prägt dies die moderne Arbeitsweise.

Edge Computing ermöglicht die Verarbeitung von Daten nahe der Quelle, was die Latenzzeit reduziert und besonders wichtig für Anwendungen wie das Internet der Dinge und Echtzeit-Analysen ist. Durch die direkte Datenverarbeitung an oder in der Nähe der Datenquelle werden Verzögerungen minimiert und diese Effizienz maximiert.

Nachhaltige IT-Ansätze konzentrieren sich auf umweltfreundliche Praktiken und die Reduzierung des ökologischen Fußabdrucks, umfassen energieeffiziente Rechenzentren und umweltbewusste IT-Entscheidungen bei der Implementierung energieeffizienter Hardware und grüner Rechenzentren spielen dabei eine entscheidende Rolle.

Metaverse repräsentiert die Vision eines gemeinsamen Online-Universums, in dem Virtual Reality, die in der realen Welt miteinander verschmelzen.

Human Augmentation ist eine Erweiterung körperlicher und kognitiver Fähigkeiten durch Technologie hat das Potenzial, die menschliche Erfahrung auf ein neues Level zu heben.

Personalisierte und prädiktive Medizin ist eine moderne IT im Gesundheitswesen und ermöglicht Behandlungen, die auf die genetische Ausstattung des Einzelnen abgestimmt sind, was eine personalisierte und prädiktive Medizin ermöglicht.

Software 2.0 mit KI schreibt mithilfe von neuronalen Netzwerken und Machine Learning Programme, was eine innovative Weiterentwicklung der Softwareentwicklung darstellt.

Web3 ist eine neue Generation des Internets, Web3, verspricht mehr Transparenz, Nutzerfreundlichkeit und Unabhängigkeit und repräsentiert somit einen bedeutenden Schritt in der evolutionären Entwicklung des World Wide Web zu gehen.^{188 189}

Die rasante Entwicklung von Informationstechnologien hat bereits heute weitreichende Auswirkungen auf verschiedenste Branchen und somit auch unsere Gesellschaft enorm verändern wird. Diese IT-Trends sind nur die Spitze des Eisbergs einer Fülle neuer Technologien, denn diese bieten aufregende Möglichkeiten Neues zu entwickeln und einzusetzen, stellen uns aber auch vor ethische und praktische Herausforderungen, die es im Zuge der fortlaufenden Digitalisierung zu meistern gilt.^{190 191}

vgl.¹⁸⁸ (Klasen, 2023)

vgl.¹⁸⁹ (Consulting, 2023)

vgl.¹⁹⁰ (Klasen, 2023)

vgl.¹⁹¹ (Consulting, 2023)

Es steht fest: Unsere Lebensweise wird maßgeblich von den IT-Trends der Zukunft beeinflusst werden – es ist an der Zeit, sich darauf vorzubereiten.

4. Rückblick und Schlussbetrachtung

Im Vorwort, dem **ersten Kapitel**, wurden die Motivation, die Methodik, Zielsetzung und die Abgrenzung sowie der Aufbau dieser Masterarbeit beschrieben.

Im **zweiten Kapitel** erfolgte eine umfassende Auseinandersetzung mit den zentralen **Komponenten der Prevention – Detection – Response**. Die Abgrenzung bezieht sich auf Detection hinsichtlich Managed SIEM sowie Security Monitoring. Hierbei wurden ausgewählte Tools wie Splunk, IBM QRadar, ManageEngine, SolarWinds, LogRhythm und Logpoint mit ihren jeweiligen Funktionen, Vor- und Nachteilen, Anwendungsbeispiele sowie vergleichenden KI-Aspekten dargelegt.

Das **dritten Kapitel** widmet sich den **Anwendungen der Künstlichen Intelligenz (KI)** mit dem Schwerpunkt auf SIEM-Monitoring und Netzwerksysteme. Es werden die wichtigsten Grundlagen der KI erläutert, mit KI-Use Cases Beispielen und den aktuellen Trends in der Informatik. Immer mit dem Ausblick auf Verbesserung der KI und der Überwachung und Sicherheit dieser neuen Technologie, die in allen Bereichen der Gesellschaft immer mehr Einfluss nimmt. Die EU hat am 13.03.2024 die erste **KI-Verordnung** mit Mehrheit im EU-Parlament zugestimmt und trat am 01. August 2024 in Kraft (24.07.2024 wurde die Final-Version veröffentlicht), um verbindliche Richtlinien für den Einsatz von KI in Unternehmen und Behörden festzulegen und die Umsetzung dieser Richtlinien in sämtlichen Bereichen bis zum Jahr 2026 umzusetzen. Basierend auf der Risikoanalyse der SOC-AI und AI-Act EU (im Anhang I) wurden ausgewählte Bausteine für das KI-Monitoring entwickelt und beschrieben, die den Anforderungen des BSI-IT-Grundschutz-Schemas annähernd entsprechen könnten (Anhang II). Diese Bausteine sind speziell darauf ausgelegt, die Sicherheit und Zuverlässigkeit von KI-Systemen zu gewährleisten. Weitere Betrachtungen können zusätzlich herangezogen werden.

Die **zukünftige Entwicklung der Künstlichen Intelligenz (KI)** im Bereich der Überwachung (Monitoring) wird maßgeblich von den zentralen Komponenten der Prevention – Detection – Response beeinflusst. Diese Komponenten bilden eine essenzielle Grundlage für die Gestaltung und Implementierung von effektiven KI-Monitoring-Lösungen, die darauf abzielen, eine proaktive und adaptive Sicherheitsstrategie zu etablieren.

Prävention bildet den ersten Eckpfeiler und setzt darauf, potenzielle Bedrohungen bereits im Vorfeld zu erkennen und zu eliminieren. Im Kontext der KI-Monitoring-Entwicklung bedeutet dies, fortschrittliche Algorithmen und Modelle zu implementieren, die in der Lage sind, Muster und Anomalien in Echtzeit zu identifizieren. Die Präventionskomponente sollte dynamisch genug sein, um sich kontinuierlich an sich ändernde Bedrohungslandschaften anzupassen. Dazu gehören beispielsweise Machine Learning-Modelle, die auf historischen Daten basieren und ständig aktualisiert werden, um aktuelle Bedrohungen zu erkennen.

Die **Detection** spielt eine entscheidende Rolle, indem sie nicht nur auf präventive Maßnahmen setzt, sondern auch darauf vorbereitet ist, ungewöhnliche Aktivitäten zu identifizieren. Im Bereich der KI-Monitoring-Systeme bedeutet dies den Einsatz von fortgeschrittenen Analysewerkzeugen, die verdächtiges Verhalten und potenzielle Sicherheitsverletzungen frühzeitig erkennen können. Hier kommen KI-gestützte Anomalieerkennung und Verhaltensanalyse zum Tragen, um Abweichungen von normalem Verhalten zu identifizieren und entsprechende Warnungen auszulösen.

Die **Response** ist darauf ausgerichtet, auf erkannte Bedrohungen schnell und effektiv zu reagieren. Im Zusammenhang mit KI-Monitoring bedeutet dies, dass automatisierte Reaktionen implementiert werden können, um Angriffe abzuwehren oder einzudämmen. Dies könnte beispielsweise die automatische Isolierung von infizierten Systemen oder die Anpassung von Sicherheitsrichtlinien in Echtzeit umfassen. Eine effiziente Reaktion erfordert die Integration von KI-Systemen mit anderen Sicherheitsinfrastrukturen.

Die **NIS2-Richtlinien**, die EU-Gesetze für Cybersicherheit beinhalten, die Widerstandsfähigkeit von Netz- und Informationssystemen gegenüber Cybersicherheitsrisiken zu stärken, siehe unter¹⁹².

In der **Zukunft wird die KI-Monitoring-Entwicklung** verstärkt auf die nahtlose Integration dieser zentralen Komponenten setzen. Das Zusammenspiel von präventiven, detektiven und reaktiven Ansätzen, gestützt durch KI, ermöglicht eine effiziente und adaptive Verteidigung gegenüber immer komplexer werdenden Cyberbedrohungen. Dabei wird die KI nicht nur als passive Beobachterin dienen, sondern aktiv dazu beitragen, Sicherheitsinfrastrukturen zu stärken und die Reaktionszeiten auf Bedrohungen zu minimieren. Die Integration von KI in Monitoring-Systemen birgt zweifellos zahlreiche Vorteile, darunter eine verbesserte Effizienz, präzisere Datenauswertung und frühzeitige Erkennung von potenziellen Problemen. Dennoch ist es unerlässlich, potenzielle Gefahren und Herausforderungen im Blick behalten, um verantwortungsbewusst und ethisch diese Technologie in Ihren Anwendungen sicherzustellen. Dabei fehlt wichtige Transparenz und Erklärbarkeit in KI-Modellen beim IT-Team, um wichtige komplexe KI-Überwachungssysteme weiterzuentwickeln. Oft agieren diese als "Black Boxes", was bedeutet, dass ihre Entscheidungsprozesse für Menschen schwer nachvollziehbar sind. Dies birgt das Risiko von Fehlinterpretationen und erschwert die Haftungs-zuweisung im Falle von unerwünschten Ergebnissen. Ein weiterer bedeutender Aspekt ist der potenzielle Einfluss von Bias und Diskriminierung in KI-Modellen. Da diese aus Trainingsdaten lernen, besteht die Gefahr, dass Vorurteile aus den Daten übernommen werden. In der Überwachung könnte dies zu systematischen Vererrungen führen, wodurch bestimmte Gruppen oder Ereignisse fälschlicherweise bevorzugt oder benachteiligt werden. Sicherheitslücken und Anfälligkeiten stellen eine ernsthafte Bedrohung dar. KI-Systeme könnten Ziel von Hackerangriffen werden, die darauf abzielen, die Modelle zu täuschen oder zu manipulieren. In sicherheitskritischen Umgebungen könnte dies zu schwerwiegenden Fehlfunktionen und falschen Sicherheitsmeldungen führen. Datenschutz und Überwachung sind weitere kritische Punkte, insbesondere wenn KI-Monitoring den Zugriff auf große Mengen sensibler Daten erfordert. Unzureichender Datenschutz könnte zu Datenschutzverletzungen und einem potenziellen Missbrauch von Informationen führen, wodurch persönliche Freiheiten beeinträchtigt werden könnten. Die Abhängigkeit von Datenqualität ist entscheidend für die Leistung von KI-Modellen. Verzerrte, unvollständige oder fehlerhafte Daten können zu fehlerhaften Schlussfolgerungen und unzuverlässigen Überwachungsergebnissen führen, was die Effektivität der KI-basierten Monitoring-Systeme erheblich beeinträchtigt. Zusätzlich zu diesen technologischen Herausforderungen gibt es rechtliche und ethische Unsicherheiten im Zusammenhang mit dem Einsatz von KI im Monitoring. Klar definierte Richtlinien sind notwendig, um Unsicherheiten bezüglich der rechtlichen Verantwortlichkeit und ethischen Standards zu vermeiden, und um möglichen Missbrauch sowie unerwünschte Auswirkungen auf individuelle Rechte zu verhindern. Hierbei kommen die unterschiedlichen Richtlinien, wie AI-Act EU, NIS2, CERT, NIST und weitere zum Einsatz, die ständig den weiteren Entwicklungen von KI-Systeme angepasst werden müssen. Insgesamt erfordert die erfolgreiche Integration von KI in Monitoring-Systemen eine ausgewogene Berücksichtigung technologischer Fortschritte und eine sorgfältige Aufmerksamkeit für ethische, rechtliche und soziale Implikationen. Nur durch eine umfassende Herangehensweise können potenzielle Gefahren minimiert und die positiven Auswirkungen dieser innovativen Technologie maximiert werden. Die Abwägung von Vorteilen und Risiken ist entscheidend, insbesondere wenn es um Technologien geht, die einen so tiefgreifenden Einfluss auf verschiedene Aspekte unserer Gesellschaft haben können. Es ist wichtig zu betonen, dass die Berücksichtigung ethischer Prinzipien und die Einhaltung rechtlicher Standards von zentraler Bedeutung sind, um einen verantwortungsbewussten Einsatz von KI zu gewährleisten. Die Transparenz von KI-Systemen, die Minimierung von Bias, Datenschutz und Datensicherheit sind einige Schlüsselfaktoren, um das Vertrauen der Öffentlichkeit zu gewinnen und mögliche negative Auswirkungen zu begrenzen. Gleichzeitig ist die Notwendigkeit von klaren rechtlichen Rahmenbedingungen und ethischen Leitlinien durch die EU und die G20-Staaten und weltweit unerlässlich, um einen angemessenen Schutz vor Missbrauch und unerwünschten

vgl.¹⁹² (Union, 2024)

Folgen gemeinsam sicherzustellen. Die Zusammenarbeit auf internationaler Ebene, um Standards und Best Practices zu etablieren, ist ebenfalls von großer Bedeutung, da Technologien wie KI keine nationalen Grenzen respektieren. Im Bereich der Cybersecurity müssen Organisationen und Regierungen proaktiv auf die sich ständig weiterentwickelnden Bedrohungen reagieren. Dies umfasst nicht nur die Entwicklung robuster Abwehrmechanismen, sondern auch die Förderung von Cyberhygiene, Schulungen für Endbenutzer und die Zusammenarbeit zwischen verschiedenen Akteuren im Bereich der Cybersicherheit. Die Betonung auf ethischer und verantwortungsbewusster Nutzung von KI sowie die Schaffung eines umfassenden rechtlichen Rahmens sind entscheidende Schritte, um die positiven Auswirkungen von KI zu maximieren und gleichzeitig Risiken zu minimieren. Die moderne Gesellschaft ist auf leistungsstarke Technologien wie KI und Quantentechnologien angewiesen, die nicht nur sicher, sondern auch transparent und stabil sind. Diese Technologien spielen eine entscheidende Rolle bei der Gestaltung und Sicherung des zukünftigen Fortschritts.¹⁹³

Um das Erlernte auch direkt umzusetzen, werde ich mich weiterhin intensiv mit diesem Thema auseinandersetzen und entwickle derzeit eine KI, die in Systemen KI-Technologien erkennt, ihre Algorithmen und Modelle im Bereich des Machine Learnings (ML) und des Deep Learnings (DL) analysiert sowie die zugehörigen Datentrainingsprozesse bewertet. Die KI soll automatisch eine Risikoeinstufung gemäß den NIST-Richtlinien und dem AI-Act EU vornehmen. Dies erfordert eine kontinuierliche Weiterentwicklung, um stets auf dem neuesten Stand der Technik zu bleiben und zukünftige Technologien, wie beispielsweise Quanten-Computing, in die Analyse mit einzubeziehen.

Hinweis: Bei dieser Masterarbeit wurde eine Plagiatsprüfung mit Scribbr¹⁹⁴ und GoThesis¹⁹⁵ Plagiatsprüfung inkl. KI-Erkennung durchgeführt.

vgl.¹⁹³ (csialtd.com.au)

¹⁹⁴ (Scribbr)

¹⁹⁵ (GoThesis)

Anhang I Risikoanalyse KI

Die Excel-Files **SOC-AI.xlsx** und **Act-EU.xlsx** werden separat zur Masterarbeit mitgeliefert, da eine Erstellung einer PDF-Datei nicht sinnvoll ist.

Anhang II KI-IT-Grundschutz-Bausteine nach BSI Schema

Die Beschreibung der Bausteine - KI-Eigenschaften wurden mit der subjektiven Betrachtung anlehnend gemäss Fraunhofer IAIS¹⁹⁶ in zusammengefasster Form erstellt: (Bewertung wurde erstellt: grundlegende Funktionalität, vorgesehener Einsatzkontext, Struktur der KI-Anwendung mit erstellte Massnahmen für Daten, KI-Komponente, Einbettung und für den Betrieb sowie Risikobewertung (niedrig, mittel, hoch, sehr hoch)).

1. Verantwortlichkeit (Accountability):

1.1. MUSS: Menschen und Organisationen müssen jederzeit die Verantwortung für die Entscheidungen und Ergebnisse eines KI-Systems übernehmen. Es muss eine klare Zuordnung der Verantwortlichkeiten für die Überwachung und Steuerung der KI-Systeme geben.

Grundlegende Funktionalität: Die KI muss so entwickelt werden, dass ihre Entscheidungslogik transparent bleibt und menschliche Eingriffe oder Überprüfungen zu jeder Zeit möglich sind. Ein Audit-Log sollte verfügbar sein, der alle Entscheidungsprozesse der KI nachvollziehbar aufzeichnet.

Vorgesehener Einsatzkontext: Die KI sollte nur in Kontexten eingesetzt werden, wo die Auswirkungen der Entscheidungen klar definiert und vorhersehbar sind und bei Anwendungen mit hohem Einfluss (z. B. medizinische Diagnosen, autonome Fahrzeuge) ist eine enge menschliche Überwachung notwendig.

Struktur der KI-Anwendung: Die Verantwortungsstruktur muss klar definiert sein, z. B. durch eine klare Hierarchie der Verantwortlichen oder einen Mechanismus, der den Einfluss von Entscheidungen auf verschiedene Abteilungen oder Personen nachverfolgt.

Maßnahmen:

Daten: Implementierung eines Monitoring-Systems zur Überwachung der Qualität der verwendeten Daten. Datensätze müssen auf Bias und ethische Fragestellungen überprüft werden.

KI-Komponente: Sicherstellen, dass alle Entscheidungen der KI dokumentiert und auditierbar sind und transparente Modelle (wie Entscheidungsbäume oder regelbasierte Systeme) verwenden.

Einbettung: Klare Verantwortlichkeitsprotokolle definieren, die auf jeder Organisationsebene überprüfbar sind.

Betrieb: Regelmäßige Audits und Inspektionen der KI-Entscheidungen, insbesondere bei kritischen Anwendungen.

Risikobewertung: Hoch, da fehlerhafte Entscheidungen könnten gravierende Auswirkungen haben, vor allem in sicherheitskritischen Bereichen und es besteht das Risiko, dass Verantwortliche ihre Pflichten nicht ausreichend wahrnehmen.

1.2. SOLLTE: Es sollte eine Möglichkeit bestehen, die Entscheidungen und Handlungen des KI-Systems nachvollziehbar zu machen, damit eine angemessene Rechenschaftspflicht gewährleistet werden kann.

Grundlegende Funktionalität: Das System muss so gestaltet sein, dass Entscheidungsprozesse nachvollziehbar sind sowie dazu sollten Erklärbarkeits-Tools integriert werden, die die Grundlage der KI-Entscheidungen erklären können.

vgl.¹⁹⁶ (IAIS, 2021)

Vorgesehener Einsatzkontext: Die KI sollte in Bereichen eingesetzt werden, in denen es möglich ist, Entscheidungen zu validieren und zu interpretieren. In undurchsichtigen, hochkomplexen Bereichen könnte dies schwierig werden.

Struktur der KI-Anwendung: Eine Schicht zur Erklärung der Entscheidungen sollte in die Struktur eingebaut werden. Diese Schicht ermöglicht es, die Entscheidungsgrundlage der KI gegenüber Menschen zu erklären.

Maßnahmen:

Daten: Sicherstellen, dass der Datensatz verständlich aufbereitet wird, um die Nachvollziehbarkeit der Entscheidungen zu gewährleisten.

KI-Komponente: Implementierung von Explainable AI (XAI)-Techniken, die die Entscheidungsfindung transparent machen.

Einbettung: Schulung des Personals, um die Erklärungen der KI verstehen und hinterfragen zu können.

Betrieb: Regelmäßige Überprüfungen der Entscheidungsprozesse der KI, um sicherzustellen, dass diese stets nachvollziehbar sind.

Risikobewertung: **Mittel**, da die Entscheidungen nicht nachvollziehbar sind, könnte das Vertrauen in die KI sinken und falsche Entscheidungen nicht bemerkt werden.

1.3. DARF: Verantwortlichkeit darf an Dritte übertragen werden, wenn eine klare und nachvollziehbare Struktur der Rechenschaftspflicht etabliert ist.

Grundlegende Funktionalität: Das System muss die Verantwortungsübergabe an externe Parteien unterstützen, indem es klare Mechanismen zur Überwachung und Steuerung bietet.

Vorgesehener Einsatzkontext: Geeignet für Szenarien, in denen externe Dienstleister beteiligt sind, etwa bei der Implementierung von KI in IT-Services. Dabei muss sichergestellt werden, dass die Kontrollmechanismen auch bei den Dritten greifen.

Struktur der KI-Anwendung: Es muss ein Vertrag oder eine Vereinbarung existieren, die festlegt, wer für welche Teile der KI verantwortlich ist und wie deren Überprüfung erfolgt.

Maßnahmen:

Daten: Klare Regeln für die Datenverarbeitung durch Dritte festlegen, um den Datenschutz zu gewährleisten.

KI-Komponente: Sicherstellen, dass externe Parteien auf die zugrunde liegende KI-Technologie geschult und für deren Handhabung verantwortlich gemacht werden können.

Einbettung: Implementierung einer nachprüfbaren Rechenschaftsstruktur, in der externe Parteien regelmäßig Bericht erstatten müssen.

Betrieb: Regelmäßige Audits und Berichte von Drittanbietern, um sicherzustellen, dass sie ihre Verantwortlichkeiten korrekt wahrnehmen.

Risikobewertung: **Mittel**, da die Übertragung der Verantwortlichkeit an Dritte birgt das Risiko, dass die Kontrolle verloren geht oder dass Dritte nicht in der Lage sind, angemessen zu reagieren.

1.4. DARF NICHT: Verantwortung darf nicht ausschließlich auf die KI-Systeme selbst abgewälzt werden, da die finale Rechenschaft immer bei den Menschen liegt.

Grundlegende Funktionalität: Das System muss so konzipiert sein, dass Menschen immer die Kontrolle über das KI-System haben und die letztliche Verantwortung tragen. Vollautonome Entscheidungsprozesse ohne menschliche Überwachung dürfen nicht erlaubt sein.

Vorgesehener Einsatzkontext: In sicherheitskritischen Anwendungen muss immer ein Mensch in der Kontrollschleife bleiben und bei weniger kritischen Anwendungen kann die KI unter Bedingungen autonom handeln, aber immer unter menschlicher Aufsicht.

Struktur der KI-Anwendung: Es muss eine klare Schnittstelle existieren, durch die Menschen jederzeit eingreifen und die Verantwortung übernehmen können.

Maßnahmen:

Daten: Sicherstellen, dass die Datengrundlage der KI von Menschen überprüft und angepasst werden kann.

KI-Komponente: Eingebaute Mechanismen zur menschlichen Kontrolle und Notabschaltung bei kritischen Entscheidungen.

Einbettung: Protokolle für menschliche Eingriffe festlegen, um sicherzustellen, dass diese jederzeit möglich sind.

Betrieb: Permanente Überwachung durch Menschen, insbesondere bei sicherheitskritischen Anwendungen, um sicherzustellen, dass die KI nicht ohne menschliche Kontrolle agiert.

Risikobewertung: Sehr hoch, da ein KI-System, das ohne menschliche Verantwortlichkeit arbeitet, kann zu gravierenden Fehlentscheidungen führen und das Risiko ist besonders in sicherheitsrelevanten Bereichen erheblich.

2. Genauigkeit (Accuracy):

2.1. MUSS: KI-Systeme müssen Ergebnisse mit hoher Genauigkeit liefern, die in Übereinstimmung mit den festgelegten Spezifikationen und realen Erwartungen stehen.

Grundlegende Funktionalität: Das System muss präzise Ergebnisse liefern, die mit den festgelegten Spezifikationen übereinstimmen.

Einsatzkontext: Kritische Anwendungen wie Medizin, autonome Systeme oder Finanzanalysen erfordern sehr hohe Genauigkeit.

Struktur der KI-Anwendung: Die KI basiert auf komplexen Modellen, die sich aus großen Datensätzen speisen und kontinuierlich verbessert werden müssen.

Maßnahmen:

Daten: Verwendung hochwertiger, annotierter Daten. Regelmäßige Datenbereinigung und kontinuierliche Aktualisierung der Trainingsdaten, um Verfälschungen zu minimieren.

KI-Komponente: Einsatz von Algorithmen, die nachweislich eine hohe Präzision bieten, einschließlich technischer Metriken wie Präzision, Recall und F1-Score. Modelle müssen regelmäßig validiert werden.

Einbettung: Integration der KI-Modelle in Systeme mit Validierungsschichten, um Ergebnisse zu überprüfen, bevor sie weiterverwendet werden.

Betrieb: Implementierung von Mechanismen zur Überwachung der Genauigkeit im laufenden Betrieb, insbesondere bei Aktualisierungen der KI-Modelle.

Risikobewertung: Hoch, da ungenaue Ergebnisse zu kritischen Fehlern führen könnten, insbesondere in sensiblen Bereichen wie Medizin oder autonomem Fahren.

2.2. SOLLTE: Die Genauigkeit der KI-Modelle sollte regelmäßig durch Validierungs- und Testprozesse überprüft und verbessert werden.

Grundlegende Funktionalität: Regelmäßige Validierung und Tests sichern langfristig hohe Genauigkeit.

Einsatzkontext: Häufig aktualisierte oder volatile Kontexte, z.B. Wettervorhersage, erfordern regelmäßige Überprüfungen der Modelle.

Struktur der KI-Anwendung: Kontinuierliches Lernen oder Modell-Updates könnten notwendig sein.

Maßnahmen:

Daten: Implementierung einer Datenpipeline, die regelmäßig aktualisierte Datensätze integriert und veraltete Daten identifiziert.

KI-Komponente: Automatisierte Tests und Validierung der KI-Modelle nach festen Zeitintervallen oder nach Modell-Updates. Nutzung von Benchmarks und Vergleich mit bestehenden Systemen.

Einbettung: Etablierung eines Frameworks, das automatisierte Tests auf Produktionssystemen durchführt, ohne den laufenden Betrieb zu unterbrechen.

Betrieb: Festlegung von Wartungs- und Testzyklen, die auf Modellverhalten und externe Veränderungen abgestimmt sind.

Risikobewertung: Mittel, da regelmäßige Validierungen Fehler rechtzeitig erkennen können, aber bei unzureichenden Ressourcen oder mangelnder Wartung könnte die Modellgenauigkeit sinken.

2.3. DARF: KI-Systeme dürfen bei komplexen oder unvorhersehbaren Szenarien gewisse Fehlertoleranzen aufweisen, wenn diese klar kommuniziert und entsprechend berücksichtigt werden.

Grundlegende Funktionalität: Bei komplexen oder dynamischen Szenarien müssen gewisse Fehlertoleranzen zugelassen werden, um Flexibilität zu ermöglichen.

Einsatzkontext: Nicht-kritische Anwendungen, z.B. Empfehlungssysteme oder Unterhaltung, können eine höhere Fehlertoleranz akzeptieren.

Struktur der KI-Anwendung: Die Anwendung könnte auf probabilistischen Modellen basieren, die Fehler in Kauf nehmen, um ein breites Spektrum von Ergebnissen zu liefern.
Maßnahmen:

Daten: Festlegung klarer Toleranzschwellen in den Trainingsdaten. Sicherstellen, dass die Datenrepräsentation verschiedene Szenarien umfasst, einschließlich der komplexen Fälle.

KI-Komponente: Entwicklung und Dokumentation klarer Fehlergrenzen für jede Ausgabe. Das System muss Fehlertoleranzen adaptiv handhaben können.

Einbettung: Fehlertoleranzen müssen durch Validierungsschichten oder Benutzerinteraktionen abgefangen werden, um kritische Auswirkungen zu vermeiden.

Betrieb: Implementierung von Alarmmechanismen, die den Betreibern signalisieren, wenn Fehlertoleranzen überschritten werden.

Risikobewertung: Niedrig bis Mittel, abhängig von der Anwendung. Bei nicht-kritischen Systemen ist das Risiko geringer, während Fehlertoleranzen in sicherheitskritischen Systemen strenger gehandhabt werden müssen.

2.4. DARF NICHT: Ergebnisse, die eine signifikante Ungenauigkeit aufweisen und negative Auswirkungen haben könnten, dürfen nicht ungeprüft bleiben.

Grundlegende Funktionalität: Systeme müssen sicherstellen, dass signifikante Fehler erkannt und behoben werden, bevor sie negative Auswirkungen haben.

Einsatzkontext: In risikoreichen Kontexten, wie z.B. medizinische Diagnostik oder autonomes Fahren, dürfen keine signifikanten Ungenauigkeiten auftreten.

Struktur der KI-Anwendung: Das System muss in der Lage sein, potenziell fehlerhafte Ergebnisse zu erkennen und zu blockieren.

Maßnahmen:

Daten: Regelmäßige Überprüfung und Validierung der Trainings- und Testdaten, um sicherzustellen, dass keine signifikanten Fehler in den Modellen auftreten.

KI-Komponente: Implementierung von Mechanismen zur Erkennung und Abmilderung von gravierenden Ungenauigkeiten, wie z.B. Outlier-Detection-Algorithmen oder Feedback-Schleifen.

Einbettung: Entwicklung von Kontrollmechanismen, die sicherstellen, dass das System automatisch bei potenziell fehlerhaften Ergebnissen eingreift.

Betrieb: Implementierung eines Eskalationsprozesses, der sicherstellt, dass bei signifikanten Fehlern menschliche Kontrolle erfolgt.

Risikobewertung: Sehr hoch, da signifikante Ungenauigkeiten in kritischen Anwendungen zu schwerwiegenden Folgen führen können, wie z.B. rechtliche oder sicherheitsrelevante Probleme.

3. Erklärbarkeit (Explainability):

3.1. MUSS: Die Entscheidungen und Empfehlungen eines KI-Systems müssen so erklärt werden können, dass sie für den Menschen verständlich sind.

Grundlegende Funktionalität: Sicherstellen, dass Entscheidungen und Empfehlungen des KI-Systems für Menschen verständlich und nachvollziehbar sind.

Vorgesehener Einsatzkontext: Einsatz in Bereichen, in denen die Nutzer genaue Erklärungen für KI-Entscheidungen benötigen, wie Gesundheitswesen, Finanzen oder Recht.

Struktur der KI-Anwendung: Eine modulare und transparente Architektur, die es ermöglicht, jede Komponente und ihren Entscheidungsprozess verständlich zu erklären.

Maßnahmen:

Daten: Auswahl von gut dokumentierten und interpretierten Datenquellen, die semantisch sinnvoll sind.

KI-Komponente: Verwendung von Algorithmen, die von Natur aus erklärbar sind (z.B. Entscheidungsbäume, regelbasierte Modelle) oder Integration von Methoden wie SHAP/LIME für komplexere Modelle.

Einbettung: Bereitstellung einer Benutzeroberfläche, die klare und verständliche Visualisierungen für die Entscheidungen bietet.

Betrieb: Regelmäßige Audits und Überprüfungen der Entscheidungen durch menschliche Experten, um sicherzustellen, dass die Erklärungen der Realität entsprechen.

Risikobewertung: Mittel, da viele erklärbar Methoden existieren, kann bei komplexen Modellen (wie neuronalen Netzen) die Erklärbarkeit schwer umsetzbar sein.

3.2. SOLLTE: Die Erklärungsmethoden sollten sowohl für technische als auch nicht-technische Nutzer angepasst werden können, um ein breiteres Verständnis zu ermöglichen.

Grundlegende Funktionalität: Ermöglichung von maßgeschneiderten Erklärungen, die sowohl für technische als auch nicht-technische Nutzer verständlich sind.

Vorgesehener Einsatzkontext: Einsatz in Umgebungen mit einer gemischten Zielgruppe, wie Unternehmen mit technischem und nicht-technischem Personal.

Struktur der KI-Anwendung: Mehrschichtige Erklärungsarchitektur, die sowohl einfache als auch detaillierte Erklärungen auf verschiedenen Abstraktionsebenen bietet.

Maßnahmen:

Daten: Daten sollten sowohl einfache als auch tiefgehende, technische Erklärungen ermöglichen.

KI-Komponente: Anpassbare Erklärungsmethoden, die für unterschiedliche Nutzergruppen flexibel gestaltet werden können, z.B. durch SHAP oder LIME.

Einbettung: Benutzeroberfläche, die zwischen technischen und nicht-technischen Erklärungen umschalten kann, mit visuellen oder textbasierten Erklärungen je nach Zielgruppe.

Betrieb: Regelmäßige Nutzertests mit unterschiedlichen Zielgruppen, um sicherzustellen, dass die Erklärungen verständlich und korrekt sind.

Risikobewertung: Mittel, da es besteht die Herausforderung, die Erklärungen für verschiedene Zielgruppen ausgewogen und sinnvoll zu gestalten, ohne wesentliche Informationen zu verlieren.

3.3. DARF: Die Erklärbarkeit darf durch vereinfachte Darstellungen oder Metaphern unterstützt werden, solange die Kernaussage nicht verzerrt wird.

Grundlegende Funktionalität: Verwendung von Metaphern und vereinfachten Darstellungen, die helfen, komplexe Entscheidungen verständlich zu machen, ohne die Kernaussage zu verzerren.

Vorgesehener Einsatzkontext: In Benutzerumgebungen, in denen schnelle und verständliche Erklärungen notwendig sind, ohne dass tiefe technische Details benötigt werden.

Struktur der KI-Anwendung: Flexibilität in der Anwendung, um sowohl vereinfachte Darstellungen als auch detaillierte technische Erklärungen für die gleiche Entscheidungslogik zu unterstützen.

Maßnahmen:

Daten: Daten sollten so aufbereitet werden, dass sie auf verschiedenen Verständnisebenen interpretiert werden können.

KI-Komponente: Modelle sollten konsistente und verständliche Entscheidungen treffen, die durch vereinfachte Darstellungen visualisiert werden können.

Einbettung: Integration von Tools wie Diagrammen oder Metaphern, um komplexe Entscheidungen verständlich darzustellen, ohne wesentliche Details zu verlieren.

Betrieb: Regelmäßige Überprüfung der Darstellungen, um sicherzustellen, dass die vereinfachten Erklärungen keine wichtigen Informationen verfälschen.

Risikobewertung: Niedrig, da die vereinfachten Erklärungen die Kernaussage richtig wiedergeben, besteht nur ein geringes Risiko und die Architektur muss jedoch sicherstellen, dass Details auf Anfrage verfügbar bleiben.

3.4. DARF NICHT: KI-Modelle dürfen keine Entscheidungen treffen, deren innere Logik oder Kausalzusammenhänge nicht nachvollziehbar sind.

Grundlegende Funktionalität: Vermeidung von Entscheidungen, deren innere Logik oder Kausalität nicht nachvollziehbar sind (keine Black-Box-Modelle ohne Erklärung).

Vorgesehener Einsatzkontext: Kritische Bereiche wie medizinische Diagnosen, Finanzentscheidungen oder rechtliche Bewertungen, in denen unverständliche Entscheidungen erhebliche Folgen haben könnten.

Struktur der KI-Anwendung: Verwendung transparenter Modelle oder ergänzender Erklärungsmodule, um sicherzustellen, dass jede Entscheidung vollständig nachvollziehbar ist.

Maßnahmen:

Daten: Die Daten sollten vollständig dokumentiert und nachvollziehbar sein, um Black-Box-Entscheidungen zu vermeiden.

KI-Komponente: Verwendung von erklärbaren oder hybrid-erklärbaren Modellen. Falls Black-Box-Modelle unvermeidbar sind, sollten zusätzliche Erklärungskomponenten wie SHAP/LIME implementiert werden.

Einbettung: Transparente Darstellung der Entscheidungslogik, sodass der Nutzer jeden Schritt der KI nachvollziehen kann.

Betrieb: Durchführung von Audits und Eingriffsmöglichkeiten für den Fall, dass eine Entscheidung nicht nachvollziehbar ist sowie ständige Überwachung und Analyse der Entscheidungslogik.

Risikobewertung: Hoch, da bei komplexen Modellen wie tiefen neuronalen Netzen ist es oft schwer, die innere Logik vollständig zu erklären so dass das Risiko ist besonders hoch, wenn das Modell in sicherheitskritischen Bereichen eingesetzt wird und Entscheidungen nicht nachvollziehbar sind.

4. Fairness:

4.1. MUSS: KI-Systeme müssen so gestaltet sein, dass sie keine Diskriminierung aufgrund von Geschlecht, Rasse, Alter, sozialem Status oder anderen persönlichen Merkmalen erzeugen.

Grundlegende Funktionalität: Die KI muss so konzipiert sein, dass sie fair gegenüber allen Benutzern agiert, unabhängig von Geschlecht, Rasse, Alter etc.

Vorgesehener Einsatzkontext: Dieses Prinzip gilt für alle KI-Systeme, insbesondere bei sensiblen Anwendungen (z.B. im Gesundheitswesen oder Finanzwesen).

Struktur der KI-Anwendung: KI-Modelle und Algorithmen müssen vor dem Einsatz auf mögliche diskriminierende Muster getestet und entsprechend optimiert werden.

Maßnahmen:

Daten: Sicherstellen, dass die Trainingsdaten repräsentativ für verschiedene Gruppen sind. Anonymisierung sensibler Daten, um Verzerrungen durch persönliche Merkmale zu vermeiden.

KI-Komponente: Einsatz von Algorithmen zur Erkennung und Beseitigung von Vorurteilen (Bias). Algorithmen zur fairen Verteilung von Entscheidungen entwickeln.

Einbettung: Implementierung von Monitoring-Tools, um die KI auf unfaire Ergebnisse hin zu überwachen.

Betrieb: Regelmäßige Audits und Fairness-Prüfungen während des laufenden Betriebs.

Risikobewertung: Mittel, da eine unzureichende Berücksichtigung von Diskriminierungsfaktoren kann zu systematischen Ungerechtigkeiten führen, jedoch sind die Risiken in den meisten Fällen mit geeigneten Maßnahmen beherrschbar.

4.2. SOLLTE: Fairness sollte durch kontinuierliche Überwachung von Vorurteilen im Modell und den verwendeten Daten sichergestellt werden. Weiterhin sollten bei veränderten Rahmenbedingungen und veränderten Nutzerverhalten regelmäßig angepasst werden.

Grundlegende Funktionalität: Die KI sollte regelmäßig auf Fairness überprüft werden.

Vorgesehener Einsatzkontext: Diese Maßnahme ist besonders wichtig in dynamischen Umgebungen, wo sich die Nutzergruppen oder -verhalten ändern können (z.B. bei Empfehlungssystemen, Finanzwesen, Bildungssystemen).

Struktur der KI-Anwendung: Modelle sollten so gestaltet sein, dass sie flexibel auf Monitoring und Anpassungen reagieren können.

Maßnahmen:

Daten: Implementierung von Tools zur Erkennung von Vorurteilen und zur Analyse von Datensatzrepräsentation und kontinuierliche Aktualisierung der Daten.

KI-Komponente: Integration von Mechanismen zur automatisierten Überprüfung und Anpassung von Modellen bei Änderungen in den Daten oder dem Nutzerverhalten.

Einbettung: Nutzung von Monitoring-Dashboards, die Metriken zur Fairness visualisieren und Alarme auslösen, wenn Diskriminierung droht.

Betrieb: Etablierung eines Wartungsplans für KI-Systeme, der regelmäßige Überprüfungen und Aktualisierungen sicherstellt.

Risikobewertung: Mittel, da dass Bias unbemerkt bleibt, besteht, ist aber durch die genannten Maßnahmen kontrollierbar sowie der Einsatz von Monitoring-Systemen minimiert dieses Risiko.

4.3. DARF: Bestimmte Daten dürfen verwendet werden, um die Fairness in spezifischen Kontexten zu fördern (z. B. gezielte Förderprogramme), solange dies transparent und ethisch gerechtfertigt ist.

Grundlegende Funktionalität: In spezifischen Kontexten dürfen persönliche Daten zur gezielten Förderung verwendet werden.

Vorgesehener Einsatzkontext: Diese Praxis ist relevant bei Förderprogrammen, sozialen Projekten oder Maßnahmen zur Chancengleichheit.

Struktur der KI-Anwendung: KI-Systeme sollten in der Lage sein, spezifische Fördergruppen ohne Verzerrungen zu erkennen und zu adressieren.

Maßnahmen:

Daten: Sicherstellen, dass die Datenverarbeitung transparent und ethisch erfolgt. Einbindung von Datenschutzmaßnahmen, um Missbrauch von sensiblen Daten zu verhindern.

KI-Komponente: Verwenden von Modellen, die eine gerechte Behandlung bestimmter Gruppen garantieren und dabei auf fairnessorientierten Zielmetriken beruhen.

Einbettung: Transparente Dokumentation und Kommunikation darüber, wie und warum bestimmte Daten verwendet werden.

Betrieb: Ethische Überprüfung der KI-Systeme durch externe Stellen. Sicherstellen, dass die Datenverwendung dem festgelegten Zweck entspricht und keine negativen Folgen für andere Nutzergruppen hat.

Risikobewertung: Niedrig, da die Verwendung von Daten transparent und ethisch gerechtfertigt ist, bleibt das Risiko gering und eine unethische Anwendung könnte jedoch zu Vertrauensverlust und rechtlichen Konsequenzen führen.

4.4. DARF NICHT: Vorurteile in den Daten oder im Algorithmus dürfen nicht unbeachtet bleiben, insbesondere wenn sie zu sozialer Ungerechtigkeit führen könnten.

Grundlegende Funktionalität: Es ist zwingend erforderlich, dass alle Formen von Vorurteilen in der KI aktiv erkannt und beseitigt werden.

Vorgesehener Einsatzkontext: Diese Anforderung betrifft alle Kontexte, in denen KI eingesetzt wird, besonders in Bereichen mit hohem Risiko für Diskriminierung (z.B. Strafjustiz, Gesundheit, Kreditvergabe).

Struktur der KI-Anwendung: Die Modelle müssen so gestaltet sein, dass sie auf bekannte Bias überprüft und bei Bedarf angepasst werden können.

Maßnahmen:

Daten: Implementierung von Prüfmechanismen, die sicherstellen, dass Vorurteile in den Trainingsdaten erkannt und korrigiert werden sowie die Nutzung von "Fairness-Toolkits" für die Datenanalyse.

KI-Komponente: Verwendung von Algorithmen, die speziell ausgelegt sind, um Vorurteile zu minimieren (z.B. adversarial fairness).

Einbettung: Etablierung von Prozessen mit KI, die regelmäßig auf Vorurteile überprüft werden und eine schnelle Intervention ermöglicht.

Betrieb: Regelmäßige Audits durch unabhängige externe IT-Experten, um sicherzustellen, dass keine unfairen Muster entstehen oder bestehen bleiben.

Risikobewertung: Hoch, da Vorurteile unbeachtet bleiben, auf Grund zu systematischer Diskriminierung führen kann und durch regelmäßige Audits und faire Algorithmen das Risiko jedoch senken können.

5. Datenschutz (Privacy):

5.1. MUSS: KI-Systeme müssen personenbezogene Daten sicher verarbeiten und schützen, indem sie die geltenden Datenschutzgesetze wie DSGVO einhalten sowie die nicht in der DSGVO enthaltenden Daten sind in konformer Nutzung. **Grundlegende Funktionalität:**

KI-Systeme, die personenbezogene Daten verarbeiten müssen sicherstellen, dass die Verarbeitung im Einklang mit den Datenschutzgesetzen DSGVO erfolgt und das betrifft die Aspekte wie Datenerhebung, -speicherung, -verarbeitung, -weitergabe und -löschung.

Vorgesehener Einsatzkontext: Die Anwendung der KI kann je nach Branche und Einsatzbereich unterschiedlich ausfallen, wie z.B. Medizin, Finanzen, Marketing. Wobei auch je Sensibilisierung der Daten und der Kontext müssen die Maßnahmen desto strenger sein.

Struktur der KI-Anwendung: Die KI-Systeme können unterschiedliche Struktur haben, wie zentrale, dezentrale oder hybride Architekturen und zudem kann die Art des Trainings und der Modellnutzung (z.B. cloudbasiert, On-Premise) einen Einfluss auf den Datenschutz haben.

Maßnahmen:

Daten: Datenverschlüsselung, die in Ruhe und bei der Übertragung, Pseudonymisierung und Anonymisierung von personenbezogenen Daten erhoben werden.

KI-Komponente: Implementierung von Mechanismen zur Löschung und Aktualisierung von Daten, um das "Recht auf Vergessenwerden" zu unterstützen.

Einbettung: Integration von Zugriffskontrollen und Authentifizierungsmechanismen in die KI-Plattform, um unbefugten Zugriff auf personenbezogene Daten zu verhindern.

Betrieb: Regelmäßige Sicherheitsprüfungen und Audits zur Sicherstellung der Einhaltung von Datenschutzbestimmungen.

Risikobewertung: **Mittel**, da diese Maßnahmen nicht korrekt implementiert werden, besteht das Risiko von Datenschutzverletzungen, was zu erheblichen rechtlichen und finanziellen Konsequenzen führen könnte.

5.2. SOLLTE: Der Zugriff auf persönliche Daten sollte auf ein Minimum beschränkt werden, und Daten sollten so weit wie möglich anonymisiert werden. **Grundlegende Funktionalität:**

Minimierung des Zugriffs auf personenbezogene Daten und Maximale Anonymisierung der Daten, um Missbrauch zu verhindern.

Vorgesehener Einsatzkontext: In allen KI-Systemen, die mit personenbezogenen Daten arbeiten (z. B. medizinische Systeme, Finanzanwendungen).

Struktur der KI-Anwendung: Zugriff auf sensible Daten sollte rollenbasiert und auf die unbedingt notwendigen Parteien beschränkt sein sowie KI-Anwendung sollte Mechanismen zur automatisierten Anonymisierung implementieren.

Maßnahmen:

Daten: Implementierung von Daten-Masking und Anonymisierungsverfahren und Reduzierung des Datenzugriffs auf das notwendige Minimum durch strenge Zugriffsrichtlinien.

KI-Komponente: Sicherstellung, dass KI-Systeme nur die erforderlichen Daten verarbeiten. Implementierung von Algorithmen zur Datenanonymisierung und -pseudonymisierung.

Einbettung: Einbettung in Systeme, die sicherstellen, dass nur autorisierte Nutzer Zugang zu personenbezogenen Daten erhalten und automatisierte Logs, die den Zugriff auf personenbezogene Daten dokumentieren.

Betrieb: Zugriffskontrollen sollten regelmäßig überprüft werden.

Schulung der Mitarbeiter in Bezug auf Datenschutz und Minimierung des Datenzugriffs.

Risikobewertung: **Niedrig**, da bei strikter Minimierung des Zugriffs und Anonymisierung der Daten ist das Risiko einer Datenschutzverletzung gering.

5.3. DARF: Daten dürfen für das Training und die Verbesserung des Systems verwendet werden, wenn die Zustimmung der Betroffenen vorliegt und der Datenschutz gewährleistet ist. **Grundlegende Funktionalität:**

Verwendung von personenbezogenen Daten zur Verbesserung der KI-Systeme unter Einhaltung des Datenschutzes.

Vorgesehener Einsatzkontext: KI-Systeme, die kontinuierlich optimiert und trainiert werden (z. B. Sprachassistenten, medizinische Diagnosesysteme).

Struktur der KI-Anwendung: Klar definierte und DSGVO-konforme Mechanismen zur Einholung der Zustimmung und zur Nachverfolgbarkeit der Nutzung personenbezogener Daten.

Maßnahmen:

Daten: Einholung und Speicherung der Einwilligung der betroffenen Personen und der Sicherstellung, dass die Verwendung der Daten auf den zugestimmten Zweck beschränkt ist.

KI-Komponente: Implementierung von Mechanismen zum Zustimmungstracking und zur Einhaltung der vereinbarten Verwendungszwecke.

Einbettung: Einbettung in eine Umgebung, die sicherstellt, dass die Zustimmung vor der Verarbeitung überprüft wird und der Bereitstellung transparenter Informationen über die Verwendung von Daten.

Betrieb: Regelmäßige Überprüfung der Einhaltung der Nutzungsbedingungen der Daten. Mechanismen zur Gewährleistung, dass Daten nach Widerruf der Zustimmung nicht weiterverwendet werden.

Risikobewertung: **Mittel**, da die Risiken entstehen durch mögliche Missverständnisse bei der Einwilligung oder durch unzureichende Schutzmaßnahmen bei der Verarbeitung der Daten.

5.4. DARF NICHT: Persönliche Daten dürfen nicht ohne ausdrückliche Zustimmung der betroffenen Person offengelegt oder missbraucht oder unerwünscht offengelegt werden.

Grundlegende Funktionalität: Sicherstellung, dass personenbezogene Daten nicht ohne Zustimmung weitergegeben oder missbraucht werden.

Vorgesehener Einsatzkontext: In allen KI-Systemen, die Zugriff auf persönliche Daten haben (z. B. personalisierte Marketing-KI, Gesundheitsanwendungen).

Struktur der KI-Anwendung: Mechanismen zur Nachverfolgbarkeit und Dokumentation des Datenflusses sowie sicherstellen, dass Daten ohne ausdrückliche Zustimmung nicht weitergegeben werden.

Maßnahmen:

Daten: Implementierung von Protokollen, die eine Weitergabe von Daten nur nach ausdrücklicher Zustimmung ermöglichen und regelmäßige Audits, um sicherzustellen, dass keine unbefugte Offenlegung stattfindet.

KI-Komponente: KI-Systeme müssen Funktionen enthalten, die eine Offenlegung ohne Zustimmung technisch verhindern (z. B. durch Verschlüsselung oder Rechtebeschränkungen).

Einbettung: Verwendung von sicheren Kommunikationskanälen und Netzwerken, um unbefugte Offenlegung zu verhindern sowie sicherstellen, dass KI-Komponenten personenbezogene Daten nur mit Zustimmung weitergeben können.

Betrieb: Regelmäßige Überprüfung der Systeme, um sicherzustellen, dass keine unerlaubten Datenlecks entstehen sowie Etablierung eines klaren Prozesses für den Widerruf von Datenverarbeitungszustimmungen.

Risikobewertung: **Hoch**, da ohne geeignete Maßnahmen besteht ein hohes Risiko für Datenschutzverletzungen, insbesondere wenn personenbezogene Daten unerlaubt weitergegeben werden.

6. Zuverlässigkeit (Reliability):

6.1. MUSS: Ein KI-System muss in der Lage sein, unter verschiedenen Bedingungen, wie fehlerfreie Eingabedaten, konsistente und verlässliche Ergebnisse zu liefern.

Grundlegende Funktionalität: Das System muss konsistente und wiederholbare Ergebnisse liefern, unabhängig von unterschiedlichen Bedingungen.

Vorgesehener Einsatzkontext: Die KI muss stabil in unterschiedlichen Szenarien funktionieren, insbesondere bei variierenden Eingabedaten.

Struktur der KI-Anwendung: Sind Modulare Architektur für die Aufteilung in einzelne Komponenten zur leichteren Wartung und Anpassung, Datenvorverarbeitungsmodul zur Sicherstellung der Datenqualität durch Vorverarbeitung, Fehlererkennungs- und Managementsystem zur Überwachung und Fehlererkennung in der Datenpipeline sowie Versionierung der Modelle in den Dokumentation und Rückverfolgbarkeit von Modelländerungen.

Maßnahmen:

Daten: Einsatz hochwertiger, gereinigter Daten und Validierung der Eingaben.

KI-Komponente: Regelmäßige Validierung der Modelle und Kalibrierung, um unter verschiedenen Bedingungen konsistent zu bleiben.

Einbettung: Echtzeitüberwachung der Leistung und Verhinderung der Verarbeitung fehlerhafter Daten.

Betrieb: Regelmäßige Audits, um sicherzustellen, dass das System stabil und zuverlässig bleibt.

Risikobewertung: **Mittel**, da Datenfehler oder ungeeignete Eingaben können zu Problemen führen, aber Maßnahmen wie Datenvalidierung und Monitoring senken das Risiko.

6.2. SOLLTE: Regelmäßige Tests und Audits sollten durchgeführt werden, um die Zuverlässigkeit des Systems zu gewährleisten, wie beispielsweise die eingebundenen Detektionsstrategien, um Eingabedaten, die nicht im Anwendungsbereich liegen herauszufiltern. Abweichungen von bei der Verarbeitung von Eingangsdaten sollten qualitativ und quantitativ betrachtet werden, um z.B. Rauschen und adversariale Verarbeitungen zu vermeiden sowie der Vermeidung von unbeabsichtigter Modellverschiebungen (Model Drifts) oder einer Veränderung des Anwendungskontexts (Concept Drift) an Leistung verlieren oder andere Anforderungen, sollten bei eingesetzte ML-Modell vermieden werden.

Grundlegende Funktionalität: Das System sollte Anomalien oder Eingaben erkennen, die nicht zum Anwendungsbereich gehören.

Vorgesehener Einsatzkontext: Eine besondere Wichtigkeit sind bei dynamischen Eingaben erforderlich, die sich über die Zeit verändern (Concept Drift, Model Drift).

Struktur der KI-Anwendung: Automatisiertes Testframework für regelmäßige Tests zur Validierung des Systems unter variierenden Bedingungen, Anomalieerkennungssystem zur Erkennung von Daten, die außerhalb des Anwendungsbereichs liegen, Monitoring- und Audit-Tools zur Echtzeitüberwachung der Ergebnisse und regelmäßige Audits, Modellüberwachungs- und -managementsystem zur Überwachung und Anpassung der Modelle bei Concept und Model Drift.

Maßnahmen:

Daten: Systeme zur Kennzeichnung und Filterung von Daten, die nicht in den Anwendungsbereich fallen.

KI-Komponente: Schutz vor adversarialen Eingaben und Mechanismen zur Erkennung von Concept Drift.

Einbettung: Tools zur kontinuierlichen Überwachung und automatischen Anpassung der Modelle.

Betrieb: Regelmäßige Audits und Tests zur Erkennung und Behebung von Abweichungen.

Risikobewertung: **Mittel bis Hoch**, da Concept Drift und adversariale Angriffe können schwerwiegende Auswirkungen haben, aber durch kontinuierliche Überwachung und Tests lässt sich das Risiko reduzieren.

6.3. DARF: Das System darf Anpassungen oder Kalibrierungen erfordern, um unter unterschiedlichen Bedingungen optimale Ergebnisse zu liefern.

Grundlegende Funktionalität: Das System bedarf Anpassungen oder Kalibrierungen erfordern, um unter unterschiedlichen Bedingungen optimal zu funktionieren. **Vorgesehener**

Einsatzkontext: Anwendungen, die flexible oder sich ändernde Eingaben verarbeiten müssen.

Struktur der KI-Anwendung: Adaptives Lernsystem zur Verwendung von kontinuierlichem oder inkrementellem Lernen zur Anpassung an neue Bedingungen, konfigurierbare Parameterkontrolle mit Tools zur einfachen Anpassung von Modellparametern und selbstkalibrierende Module zur automatische Kalibrierung basierend auf eingehenden Daten und Rollback-Mechanismus mit Möglichkeit, zu einer früheren stabilen Version zurückzukehren, falls eine Kalibrierung unerwünschte Ergebnisse liefert.

Maßnahmen:

Daten: Sammlung und Analyse von Daten aus unterschiedlichen Quellen, um Anpassungen vorzunehmen.

KI-Komponente: Modelle sollten in der Lage sein, sich flexibel an neue Bedingungen anzupassen (z.B. durch Transfer Learning).

Einbettung: Sicherstellung der korrekten Kalibrierung durch Kontrollmechanismen.

Betrieb: Tests nach jeder Anpassung oder Kalibrierung, um die Leistungsfähigkeit sicherzustellen.

Risikobewertung: Niedrig bis Mittel, da eine sorgfältige Kalibrierung minimiert das Risiko, aber falsche Anpassungen können unerwartete Ergebnisse liefern.

6.4. DARF NICHT: Ein System darf keine unvorhersehbaren oder stark variierenden Ergebnisse liefern, insbesondere wenn dies sicherheitskritische Auswirkungen haben könnte.

Grundlegende Funktionalität: Unvorhersehbare Ergebnisse müssen vermieden werden, besonders wenn sie sicherheitskritische Auswirkungen haben könnten.

Vorgesehener Einsatzkontext: Anwendungen mit hohen Anforderungen an Sicherheit und Stabilität, wie autonome Systeme oder medizinische Anwendungen.

Struktur der KI-Anwendung: Sicherheitskritische Architektur für redundante Systeme zur Sicherstellung der Zuverlässigkeit bei sicherheitskritischen Anwendungen; Verifizierte und validierte Module zur strikten Überprüfung und Validierung der sicherheitskritischen Teile der KI-Anwendung; Fehlerkorrektur- und Eskalationssystem mit automatischer Fehlererkennung und Korrekturmechanismen sowie der Erklärungsmechanismus (Explainability) für Mechanismen, die Entscheidungen des KI-Systems erklärbar und nachvollziehbar machen.

Maßnahmen:

Daten: Strenge Qualitätskontrollen, um sicherzustellen, dass fehlerhafte oder ungeeignete Daten keine variierenden Ergebnisse hervorrufen.

KI-Komponente: Modelle, die auch bei fehlerhaften Eingaben robuste und konsistente Ergebnisse liefern.

Einbettung: Überwachungssysteme zur Erkennung und Behandlung von unvorhersehbaren Ergebnissen in Echtzeit.

Betrieb: Strenge Sicherheits- und Zuverlässigkeitstests, insbesondere vor dem Einsatz in kritischen Umgebungen.

Risikobewertung: Sehr hoch, da unvorhersehbare oder variierende Ergebnisse in sicherheitskritischen Anwendungen sind inakzeptabel und können zu katastrophalen Folgen führen sowie Redundanz, strikte Überwachung und Tests senken das Risiko, aber es bleibt hoch.

7. Resilienz (Resiliency):

7.1.MUSS: Ein KI-System muss in der Lage sein, nach einem feindlichen Angriff oder technischen Ausfall schnell wieder in einen sicheren Betriebszustand zurückzukehren.

Grundlegende Funktionalität: Die Fähigkeit, schnell in den sicheren Betriebszustand zurückzukehren, erfordert Mechanismen zur Fehlererkennung und Wiederherstellung. Automatische Wiederherstellung ohne menschliches Eingreifen ist wichtig.

Vorgesehener Einsatzkontext: In sicherheitskritischen Anwendungen (z. B. autonome Fahrzeuge, medizinische Systeme) muss das System in kürzester Zeit wiederhergestellt werden sowie bei weniger kritischen Anwendungen (z. B. KI für Marketing) sind die Anforderungen möglicherweise weniger streng.

Struktur der KI-Anwendung: Die Architektur muss redundante und selbstüberwachende Mechanismen enthalten.

Maßnahmen:

Daten: Regelmäßige Sicherung und Replikation von Daten auf mehreren sicheren Servern, um Datenverlust bei einem Ausfall zu minimieren.

KI-Komponente: Einsatz von Fehlertoleranz-Mechanismen in der KI, wie z. B. Hot-Backup-Modelle oder parallele Modelle.

Einbettung: Integration von automatisierten Wiederherstellungsmechanismen im Rahmen der eingebetteten KI-Anwendung.

Betrieb: Monitoring-Systeme, die in Echtzeit den Systemstatus überwachen, zusammen mit Protokollen zur schnellen Wiederherstellung.

Risikobewertung: Mittel bis Hoch, da je nach Einsatzbereich variiert das Risiko. In sicherheitskritischen Bereichen wäre das Risiko sehr hoch, wenn Resilienz nicht gewährleistet ist.
7.2. SOLLTE: Es sollte eine proaktive Planung zur Minimierung von Ausfallzeiten und zur schnellen Wiederherstellung von Funktionen geben.

Grundlegende Funktionalität: Proaktive Planung erfordert Ausfallrisikoanalysen und Pläne zur Minimierung von Ausfallzeiten.

Vorgesehener Einsatzkontext: In Systemen mit hoher Verfügbarkeit, wie z. B. in der Finanzbranche oder in sicherheitskritischen Bereichen, ist eine proaktive Planung unerlässlich.

Struktur der KI-Anwendung: Resilienz sollte bei der Architektur von Anfang an berücksichtigt werden, durch Redundanz und Backup.

Maßnahmen:

Daten: Erstellung von Backups in Echtzeit und das regelmäßige Testen von Wiederherstellungsprozessen.

KI-Komponente: Nutzung von Modellen mit reduzierter Rechenleistung als Fallback-Option bei Systemüberlastungen oder Ausfällen.

Einbettung: Planung und Implementierung redundanter Hardware und virtueller Maschinen zur Sicherstellung eines kontinuierlichen Betriebs.

Betrieb: Regelmäßige Durchführung von Simulationen und Notfalltests, um die Effektivität von Wiederherstellungsplänen zu überprüfen.

Risikobewertung: Mittel, da abhängt von der Kritikalität der Anwendung, und in weniger kritischen Bereichen ist es moderat, aber in kritischen Systemen, wo es zu wirtschaftlichen oder physischen Schäden kommen könnte, ist es höher.

7.3. DARF: Das System darf alternative Betriebsmodi oder Sicherungen verwenden, um seine Resilienz zu stärken.

Grundlegende Funktionalität: Alternative Betriebsmodi (z. B. „Safe Mode“) können die Auswirkungen eines Ausfalls abmildern.

Vorgesehener Einsatzkontext: In Systemen, die für den kontinuierlichen Betrieb entscheidend sind (z. B. Verkehrssteuerungssysteme), müssen alternative Betriebsmodi vorhanden sein.

Struktur der KI-Anwendung: Alternative Betriebsmodi müssen in der Software- und Hardwarestruktur von Anfang an vorgesehen werden.

Maßnahmen:

Daten: Sicherstellung, dass alternative Betriebsmodi auf minimale Datensätze zugreifen und dabei immer noch sicher agieren.

KI-Komponente: Entwicklung von vereinfachten oder abgespeckten Modellen, die in Situationen aktiviert werden können, in denen die volle KI-Funktionalität nicht verfügbar ist.

Einbettung: Bereitstellung von redundanten Hardware-Subsystemen, die in den alternativen Betriebsmodus wechseln können.

Betrieb: Entwicklung eines abgestuften Protokolls, das zwischen verschiedenen Betriebsmodi nahtlos wechseln kann, ohne den Endnutzer zu gefährden.

Risikobewertung: Niedrig bis Mittel, da Alternativmodi vorhanden sind und richtig implementiert sind. Ohne solche Modi steigt das Risiko.

7.4. DARF NICHT: Ein KI-System darf nach einem Angriff oder Ausfall nicht langfristig unbrauchbar werden oder den Nutzern Schaden zufügen.

Grundlegende Funktionalität: Das System muss gegen dauerhafte Beeinträchtigungen geschützt sein, und es müssen Maßnahmen vorhanden sein, um potenzielle Schäden für den Nutzer zu verhindern.

Vorgesehener Einsatzkontext: In sicherheitsrelevanten Anwendungen (z. B. in der Medizin oder im Transport) ist dieser Punkt besonders kritisch, da hier ein Ausfall schwerwiegende Konsequenzen haben könnte.

Struktur der KI-Anwendung: Ein KI-System muss sowohl durch Software als auch durch Hardware-Schutzmechanismen vor irreversiblen Schäden geschützt sein.

Maßnahmen:

Daten: Verschlüsselte Speicherung und regelmäßige Überprüfung der Integrität der Daten, um sicherzustellen, dass nach einem Angriff keine Datenkorruption vorliegt.

KI-Komponente: Implementierung von Schutzmechanismen, die sicherstellen, dass ein Angriff das Modell nicht permanent beeinträchtigen kann, wie z. B. Rollbacks auf frühere Versionen.

Einbettung: Sicherstellung, dass die Hardware so konzipiert ist, dass sie nach einem Ausfall oder Angriff schnell wiederhergestellt werden kann.

Betrieb: Sicherheitsprotokolle, die den Betrieb des Systems in unsicheren Zuständen automatisch unterbrechen, um Schäden zu vermeiden.

Risikobewertung: **Hoch**, da das System nach einem Angriff unbrauchbar wird, ist besonders hoch in sicherheitskritischen und finanziell sensiblen Bereichen sowie diese Resilienz nicht gewährleistet ist, können irreversible Schäden eintreten.

8. Robustheit (Robustness):

8.1. MUSS: KI-Systeme müssen robust genug sein, um unter allen Umständen eine Mindestleistung zu gewährleisten, auch unter extremen Bedingungen oder Angriffen nicht manipuliert, verfälscht oder nicht verfügbar sind durch Fehlfunktionen, Ausfälle oder unbeabsichtigter Personen- oder Sachschäden, um Gefährdungen zu vermeiden.

Grundlegende Funktionalität: Sicherstellen, dass das KI-System in der Lage ist, seine primäre Aufgabe zu erfüllen, selbst wenn Daten unvollständig oder fehlerhaft sind.

Vorgesehener Einsatzkontext: Systeme, die in sicherheitskritischen oder unvorhersehbaren Umgebungen eingesetzt werden (z. B. autonomes Fahren oder medizinische Diagnosen), müssen unter extremen Bedingungen funktionieren.

Struktur der KI-Anwendung: Die Architektur der KI sollte eine Modularität aufweisen, so dass Ausfälle oder Angriffe auf einzelne Komponenten die Gesamtfunktion nicht beeinträchtigen.

Maßnahmen:

Daten: Validierung und Redundanz sicherstellen, alternative Datenquellen implementieren.

KI-Komponente: Robuste Algorithmen mit Fehlertoleranz und Selbstheilungsmechanismen verwenden.

Einbettung: Redundante Systeme und Backups für Hardware und Infrastruktur einbetten.

Betrieb: Laufende Überwachung, Sicherheitsupdates und Notfallpläne implementieren.

Risikobewertung: **Hoch**, da Ausfälle unter extremen Bedingungen schwerwiegende Folgen haben können.

8.2. SOLLTE: Robustheit sollte durch Sensitivitätsanalysen regelmäßig geprüft und bei Bedarf angepasst werden.

Grundlegende Funktionalität: Sicherstellen, dass das KI-System keine unvorhergesehenen Reaktionen auf geringfügige Änderungen in den Eingangsdaten zeigt.

Vorgesehener Einsatzkontext: In dynamischen Umgebungen, in denen sich die Datenlage häufig ändert, sind Sensitivitätsanalysen von großer Bedeutung (z. B. Finanzmärkte, dynamische Benutzerumgebungen).

Struktur der KI-Anwendung: Die KI sollte regelmäßig auf ihre Reaktion gegenüber geänderten Bedingungen getestet werden, um sicherzustellen, dass kleine Änderungen in den Eingangsdaten keine großen Fehler verursachen.

Maßnahmen:

Daten: Regelmäßige Überprüfung der Sensitivität gegenüber Datenänderungen.

KI-Komponente: Periodische Modelltests und Anpassungen bei festgestellten Schwächen.

Einbettung: Überprüfung der Hardware-Infrastruktur, um Schwachstellen zu eliminieren.

Betrieb: Sensitivitätsanalysen im Betrieb, um Anpassungen bei Bedarf vorzunehmen.

Risikobewertung: **Mittel**, da fehlende Sensitivitätsanalysen zu unerkannten Schwächen führen können.

8.3. DARF: KI-Systeme dürfen in besonderen Ausnahmefällen von der normalen Funktion abweichen, wenn diese Abweichungen im Rahmen der Erwartungen liegen.

Grundlegende Funktionalität: Das System muss in der Lage sein, unter außergewöhnlichen Bedingungen akzeptable, aber nicht perfekte Ergebnisse zu liefern.

Vorgesehener Einsatzkontext: Systeme, die in sich verändernden oder unvorhersehbaren Umgebungen eingesetzt werden (z. B. Wettervorhersage, dynamische Verkehrssteuerung), müssen in der Lage sein, begrenzte Abweichungen zu tolerieren.

Struktur der KI-Anwendung: Die Architektur muss so gestaltet sein, dass es Fail-Safe-Mechanismen gibt, die in Ausnahmefällen greifen.

Maßnahmen:

Daten: Definition und Simulation von Ausnahmefällen zur Sicherstellung der Systemreaktion.

KI-Komponente: Implementierung von Fail-Safe-Mechanismen für vorhersehbare Ausnahmen.

Einbettung: Anpassungsfähige Infrastruktur für dynamische Bedingungen.

Betrieb: Manuelle und automatische Notfallprozeduren für Ausnahmefälle.

Risikobewertung: **Niedrig**, da erwartete Abweichungen handhabbar sind.

8.4. DARF NICHT: Ein KI-System darf nicht unter minimalen Abweichungen von den üblichen Bedingungen zusammenbrechen oder versagen.

Grundlegende Funktionalität: Selbst bei kleinen Datenanomalien oder unerwarteten Inputs muss das System weiterhin stabil arbeiten.

Vorgesehener Einsatzkontext: In Anwendungen, in denen die Zuverlässigkeit von entscheidender Bedeutung ist (z. B. in der Industrie oder in der Gesundheitsversorgung), dürfen kleine Abweichungen und keine Systemfehler verursachen.

Struktur der KI-Anwendung: Die Architektur sollte robuste Fehlerhandlungsmechanismen integrieren, um kleine Abweichungen effizient zu korrigieren ohne die Gesamtfunktion zu beeinträchtigen.

Maßnahmen:

Daten: Robustheitstests bei kleinen Abweichungen, Erkennung von minimalen Datenfehlern durchführen.

KI-Komponente: Einsatz von Modellen mit hoher Fehlertoleranz bereinigen.

Einbettung: Nutzung von stabilen Infrastrukturen, die kleine Abweichungen tolerieren.

Betrieb: Frühwarnsysteme zur Überwachung und sofortigen Korrektur kleiner Fehler überprüfen.

Risikobewertung: **Mittel**, da kleine Abweichungen zu unnötigen Ausfällen führen könnten, wenn das System nicht richtig implementiert ist.

9. Sicherheit (Safety):

9.1. MUSS: Ein KI-System muss so konzipiert sein, dass es keine unbeabsichtigten Schäden für Benutzer oder die Gesellschaft verursacht.

Grundlegende Funktionalität: Das System muss sicherstellen, dass es keine Entscheidungen trifft, die zu physischen, emotionalen oder finanziellen Schäden führen können.

Vorgesehener Einsatzkontext: Allgemeine oder kritische Anwendungen, wie z.B. Medizin, autonome Systeme.

Struktur der KI-Anwendung: ML, Regelbasierte Systeme und hybride Modelle.

Maßnahmen:

Daten: Sicherstellung, dass die Trainingsdaten korrekt und repräsentativ für den Einsatzkontext sind, um Fehlentscheidungen zu minimieren.

KI-Komponente: Entwicklung einer robusten Fehlermeldungs- und Sicherheitsstrategie (z.B. Überwachung des Modells auf anomale Entscheidungen).

Einbettung: Integration eines Feedbackmechanismus, der den Benutzer über potenziell risikobehaftete Aktionen informiert.

Betrieb: Regelmäßige Audits und Tests des Systems, um sicherzustellen, dass es sich innerhalb der spezifizierten Grenzen bewegt.

Risikobewertung: **Hoch**, da insbesondere bei kritischen Systemen (z.B. Gesundheitswesen, autonome Fahrzeuge) unbeabsichtigte Schäden gravierende Folgen haben können.

9.2. SOLLTE: Es sollten Sicherheitsmaßnahmen in das Design integriert werden, die potenziell gefährliches Verhalten verhindern.

Grundlegende Funktionalität: Systeme sollten Sicherheitsvorkehrungen (z.B. Notabschaltung oder Fail-Safe-Mechanismen) haben.

Vorgesehener Einsatzkontext: Anwendungen mit erhöhtem Risiko für Benutzer (z.B. Industriesysteme, autonome Systeme).

Struktur der KI-Anwendung: Selbstlernende und autonome Systeme mit adaptivem Verhalten.

Maßnahmen:

Daten: Kontinuierliche Überprüfung der Datenquellen und Filterung bössartiger oder manipulierte Daten.

KI-Komponente: Implementierung von Überwachungsalgorithmen, die gefährliches Verhalten erkennen und Korrekturmaßnahmen einleiten.

Einbettung: Einbau von Notfallprotokollen, die automatisch ausgelöst werden, wenn das KI-System abnormales Verhalten zeigt.

Betrieb: Einrichtung einer Überwachungsinfrastruktur, um das Verhalten des KI-Systems in Echtzeit zu kontrollieren und auf gefährliche Situationen zu reagieren.

Risikobewertung: **Mittel bis hoch**, da abhängig vom Einsatzkontext und den möglichen Auswirkungen von Fehlfunktionen.

9.3. DARF: Sicherheitsstandards dürfen je nach Anwendungsfall und Risiko abgestuft sein, solange die Grundsicherheit gewährleistet ist.

Grundlegende Funktionalität: Die Sicherheitsstandards sollten dem Risiko des spezifischen Einsatzes angepasst werden, ohne die Grundsicherheit zu gefährden.

Vorgesehener Einsatzkontext: Unterschiedliche Kontexte (niedriges Risiko: Social Media-Filter, hohes Risiko: autonome Fahrzeuge).

Struktur der KI-Anwendung: Von Regelbasierten bis zu lernenden Systemen.

Maßnahmen:

Daten: Sicherstellen, dass bei risikoreicheren Anwendungen strengere Datenqualitätsstandards gelten.

KI-Komponente: Differenzierte Sicherheitsprotokolle abhängig von der Kritikalität der Anwendung (niedrigere Sicherheitsmaßnahmen für Systeme mit geringen Risiken, strengere für sicherheitskritische Anwendungen).

Einbettung: Abstufung der Notfallprotokolle, abhängig von der Kritikalität der Anwendung.

Betrieb: Implementierung von maßgeschneiderten Überwachungs- und Wartungsmaßnahmen entsprechend dem Risikoprofil der Anwendung.

Risikobewertung: **Niedrig bis hoch**, da abhängig von der Anwendung und der Risikostufe.

9.4. DARF NICHT: KI-Systeme dürfen keine sicherheitskritischen Aufgaben übernehmen, wenn nicht nachgewiesen ist, dass sie alle notwendigen Sicherheitsanforderungen erfüllen.

Grundlegende Funktionalität: Systeme müssen einen strengen Verifikationsprozess durchlaufen, bevor sie sicherheitskritische Aufgaben übernehmen.

Vorgesehener Einsatzkontext: Hochkritische Anwendungen (z.B. Medizin, Verkehr, Verteidigung).

Struktur der KI-Anwendung: Typischerweise hochkomplexe, autonome Systeme.

Maßnahmen:

Daten: Sicherstellen, dass die Trainings- und Testdaten für die sicherheitskritische Aufgabe umfassend geprüft wurden.

KI-Komponente: Durchführung intensiver Tests und Validierungsverfahren, um sicherzustellen, dass das System den sicherheitskritischen Anforderungen gerecht wird.

Einbettung: Strenge Zertifizierungsanforderungen, bevor die KI-Komponente in sicherheitskritische Infrastrukturen eingebettet wird.

Betrieb: Laufende Überprüfung und Zertifizierung, um sicherzustellen, dass das System alle sicherheitsrelevanten Vorgaben weiterhin erfüllt.

Risikobewertung: **Sehr hoch**, da unsichere Systeme in sicherheitskritischen Kontexten zu katastrophalen Folgen führen könnten.

10. Sicherheitsmaßnahmen (Security):

10.1. MUSS: KI-Systeme müssen Maßnahmen implementieren, die den Schutz vor unbefugtem Zugriff, Angriffen und Datenlecks gewährleisten.

Grundlegende Funktionalität: Schutz der KI-Systeme vor unerlaubtem Zugriff und Missbrauch ist essenziell für die Sicherstellung ihrer korrekten Funktion. Dies betrifft alle Aspekte der Datenverarbeitung und die KI-Komponenten.

Vorgesehener Einsatzkontext: Besonders kritisch, wenn sensible oder personenbezogene Daten verarbeitet werden, wie in Gesundheits-, Finanz- oder Regierungssystemen sowie in offenen Umgebungen (z.B. öffentliche KI-Systeme) wird das Risiko höher.

Struktur der KI-Anwendung: Die Anwendung besteht typischerweise aus verschiedenen Komponenten, wie Datenvorverarbeitung: Datenaufbereitung und Bereinigung, Modelltraining der Prozess des Trainings des KI-Modells, Modellinferenz der Phase der Nutzung des trainierten Modells für Vorhersagen sowie Schnittstellen als APIs oder Benutzerschnittstellen für Interaktionen.

Maßnahmen:

Daten: Einsatz von Zugangskontrollen, Datenverschlüsselung im Ruhezustand und bei der Übertragung (z. B. TLS), regelmäßige Audits der Zugriffsrechte.

KI-Komponente: Sicherstellung der Laufzeitintegrität durch Verwendung von Signaturen und Laufzeitüberwachung (z. B. Intrusion Detection).

Einbettung: Physische Sicherheitsmaßnahmen (z. B. gesicherte Serverräume), Isolation kritischer Komponenten (Sandboxing, Virtualisierung).

Betrieb: Regelmäßige Sicherheitsüberprüfungen und Audits, Sicherheits-Updates und Patches.

Risikobewertung: Hoch, da Unzureichender Schutz könnte schwerwiegende Folgen haben, wie z. B. Datenlecks, Manipulation der KI-Ergebnisse oder unbefugten Zugang zu sensiblen Informationen.

10.2. SOLLTE: Verschlüsselung und Authentifizierungsmechanismen sollten verwendet werden, um die Integrität und Vertraulichkeit der Daten zu wahren.

Grundlegende Funktionalität: Verschlüsselung und Authentifizierung gewährleisten, dass die Daten während der Verarbeitung und Übertragung nicht manipuliert oder abgefangen werden können.

Vorgesehener Einsatzkontext: Wichtiger in Umgebungen, in denen vertrauliche oder personenbezogene Daten verarbeitet werden, wie in Cloud-Diensten, IoT-Geräten oder bei der Übertragung zwischen verschiedenen Systemen.

Struktur der KI-Anwendung: Datenvorverarbeitung zum Sicherstellen, dass nur authentifizierte und verschlüsselte Daten verarbeitet werden; Modelltraining zur Verschlüsselung der Trainingsdaten und Speicherung von Modellen mit Zugriffskontrollen sowie Modellinferenz zur Verwendung von Authentifizierung für Benutzer, die auf das Modell zugreifen sowie Schnittstellen im Einsatz von TLS für API-Schnittstellen und starke Authentifizierungsmethoden (z.B. Zwei-Faktor-Authentifizierung).

Maßnahmen:

Daten: End-to-End-Verschlüsselung (z. B. AES-256), Multi-Faktor-Authentifizierung (MFA), regelmäßige Schlüsselrotation.

KI-Komponente: Verschlüsselung von Modellen und Trainingsdaten, Verschlüsselung der Kommunikation zwischen KI-Komponenten.

Einbettung: Sichere Authentifizierung zwischen Subsystemen, verschlüsselte Datenübertragung.

Betrieb: Sicherer Zugang zu Managementschnittstellen, Protokollierung und Überwachung sicherheitsrelevanter Zugriffe.

Risikobewertung: Mittel, da das Fehlen von Verschlüsselung und Authentifizierung könnte zu unerwünschtem Datenzugriff führen, ist jedoch in einigen Anwendungsfällen weniger kritisch

10.3. DARF: Sicherheitsmaßnahmen müssen an aktuelle Bedrohungen angepasst werden, um neuen Herausforderungen gerecht zu werden.

Grundlegende Funktionalität: Die Fähigkeit, Sicherheitsmaßnahmen kontinuierlich an neue Bedrohungen anzupassen, ist wichtig, um die langfristige Sicherheit der KI-Anwendung zu gewährleisten.

Vorgesehener Einsatzkontext: Besonders relevant in dynamischen Umgebungen, wie dem Internet der Dinge (IoT) oder bei Systemen, die sich in ständiger Weiterentwicklung befinden. Struktur der KI-Anwendung:

Struktur der KI-Anwendung: Datenvorverarbeitung von Daten muss regelmäßig auf neue Bedrohungen hin geprüft und gesichert werden; Modelltraining im Training und Modellaktualisierung müssen flexibel sein, um gegen neue Arten von Angriffen (z. B. adversarial attacks) geschützt zu sein; Modellinferenz in der Laufzeitumgebung muss in der Lage sein, Sicherheitsupdates und Patches dynamisch zu implementieren sowie die Schnittstellen in der Flexibilität bei der Einführung neuer Sicherheitsprotokolle und Mechanismen, um auf entdeckte Schwachstellen zu reagieren.

Maßnahmen:

Daten: Regelmäßige Überprüfung der Datenquellen und Sicherheitsanalysen, um potenzielle Schwachstellen zu identifizieren.

KI-Komponente: Anpassung von Sicherheitsalgorithmen und Modellen basierend auf neuen Bedrohungsanalysen.

Einbettung: Implementierung adaptiver Schutzmaßnahmen (z.B. dynamische Firewalls, AI-basierte Bedrohungserkennung).

Betrieb: Incident-Response-Plans, regular Schwachstellenanalysen und schnelle Bereitstellung von Sicherheits-Patches.

Risikobewertung: **Mittel**, da ohne kontinuierliche Anpassung könnten neue Bedrohungen die Sicherheit gefährden, insbesondere in Umgebungen mit langer Betriebszeit.

10.4. DARF NICHT: KI-Systeme dürfen keine Schwachstellen aufweisen, die durch mangelnde Sicherheitsvorkehrungen ausgenutzt werden können.

Grundlegende Funktionalität: Das Vermeiden von Sicherheitslücken ist unerlässlich, um das Risiko von Angriffen, Datenverlusten oder anderen Sicherheitsvorfällen zu minimieren.

Vorgesehener Einsatzkontext: Hochkritische Systeme wie im Finanzwesen, Gesundheitswesen oder Regierungsdiensten, bei denen eine Schwachstelle schwerwiegende Folgen haben könnte.

Struktur der KI-Anwendung: Datenvorverarbeitung zur Sicherheitsprüfungen sollten bereits bei der Datenaufbereitung beginnen, um potenzielle Schwachstellen zu identifizieren; der Modelltraining zum Schutz der Trainingsumgebung und -daten gegen Manipulation; der Modellinferenz zur Sicherstellung, dass keine Schwachstellen in den Vorhersagemodellen bestehen (z.B. Schutz gegen adversarial attacks) sowie Schnittstellen für alle Kommunikationswege und Schnittstellen müssen gegen bekannte Angriffe geschützt werden (z.B. SQL-Injection, XSS).

Maßnahmen:

Daten: Regelmäßige Schwachstellenanalysen, Penetrationstests und Audits der Datensicherheitsmaßnahmen.

KI-Komponente: Prüfung von Modellen auf potenzielle Schwachstellen, Implementierung robuster Schutzmaßnahmen gegen adversarial attacks.

Einbettung: Sicherstellen, dass alle verwendeten Bibliotheken und Frameworks aktuell sind und keine bekannten Schwachstellen enthalten.

Betrieb: Einführung eines strikten Schwachstellenmanagements, kontinuierliche Überwachung und schnelles Beheben entdeckter Lücken.

Risikobewertung: **Sehr hoch**, da das Vorhandensein von Schwachstellen kann schwerwiegende Auswirkungen haben, insbesondere in kritischen Umgebungen, in denen Datenlecks oder Manipulation verheerende Folgen haben könnten.

11. Transparenz (Transparency):

11.1. MUSS: Es muss möglich sein, die Entscheidungen und Funktionsweise eines KI-Systems transparent zu erklären, um Vertrauen bei den Nutzern zu schaffen besonders bei der Nachvollziehbarkeit und Auditfähigkeit sowie der Dynamik der KI-Anwendungen.

Grundlegende Funktionalität: Das KI-System muss in der Lage sein, Entscheidungen klar und nachvollziehbar zu erklären, insbesondere für kritische Anwendungen, in denen Ver-

trauen und Nachvollziehbarkeit wesentlich sind (z.B. in der Medizin, im Finanzwesen, bei sicherheitskritischen Systemen).

Vorgesehener Einsatzkontext: System wird in Bereichen eingesetzt, in denen die KI Entscheidungen trifft, die einen erheblichen Einfluss auf Menschen oder Organisationen haben (z.B. Diagnoseunterstützung, Kreditanalysen).

Struktur der KI-Anwendung: Komplexe, datengetriebene Modelle, möglicherweise neuronale Netze oder andere Black-Box-Modelle, die schwer zu interpretieren sind.

Maßnahmen:

Daten: Dokumentation der Datenquellen und der Datenverarbeitungsschritte und Sicherstellung der Datenqualität, Transparenz über den Ursprung der Daten (z.B. welche Daten für Entscheidungen herangezogen werden) und Verwendung von Audit-Trails, um nachvollziehen zu können, welche Daten welche Entscheidung beeinflusst haben.

KI-Komponente: Integration von Explainable AI (XAI)-Techniken, um den Entscheidungsprozess transparent zu machen und die Schaffung von Protokollen zur Überwachung von Entscheidungswegen, z.B. durch lokale Erklärungsmodelle wie LIME oder SHAP sowie die Bereitstellung von Informationen über das Modelltraining, die Architektur und die genutzten Algorithmen.

Einbettung: Implementierung von User Interfaces, die den Entscheidungspfad der KI visuell und einfach verständlich darstellen sowie der API-Schnittstellen, die auf Anfrage Erklärungen der Entscheidungen zurückgeben.

Betrieb: Sicherstellung, dass alle entscheidungsrelevanten Vorgänge protokolliert und nachvollziehbar bleiben und die Bereitstellung eines regelmäßigen Auditing-Mechanismus zur Validierung der Entscheidungsprozesse.

Risikobewertung: Hoch, da ohne Transparenz in den Entscheidungen kann das Vertrauen in die KI stark beeinträchtigt werden, insbesondere in sicherheitskritischen und rechtlichen Kontexten.

11.2. SOLLTE: Stakeholder sollten regelmäßig über die Funktionsweise und Grenzen des Systems informiert werden.

Grundlegende Funktionalität: Stakeholder müssen kontinuierlich und verständlich über die Funktionsweise des Systems informiert werden, um Vertrauen und korrektes Verständnis sicherzustellen.

Vorgesehener Einsatzkontext: Verwendung in Umgebungen, in denen Nutzer und Stakeholder auf die Effizienz und Zuverlässigkeit der KI angewiesen sind (z.B. industrielle Automatisierung, Kundenservice).

Struktur der KI-Anwendung: KI-System, das regelmäßig gewartet oder aktualisiert wird und in dem sich Modelle oder Algorithmen verändern können.

Maßnahmen:

Daten: Regelmäßige Berichterstattung über die Datenqualität und die Anpassung von Datenquellen sowie sicherstellen, dass Stakeholder über relevante Änderungen in den Datenquellen informiert werden (z.B. Änderungen in der Datenpolitik oder neue Datensätze).

KI-Komponente: Bereitstellung verständlicher Dokumentationen über die Funktionsweise und eventuelle Updates des KI-Systems sowie regelmäßige Schulungen und Präsentationen, um den Wissensstand der Stakeholder auf dem aktuellen Stand zu halten.

Einbettung: Entwicklung von Dashboards, die relevante Informationen zur Systemleistung, neuen Modellen oder Algorithmen für die Stakeholder bereitstellen und Implementierung von automatisierten Benachrichtigungssystemen, die Stakeholder bei Änderungen informieren.

Betrieb: Erstellung von Kommunikationsrichtlinien, die sicherstellen, dass Stakeholder regelmäßig über Aktualisierungen und wichtige Änderungen informiert werden sowie die Organisation von Feedbackrunden, um sicherzustellen, dass die Stakeholder die Informationen verstehen und anwenden können.

Risikobewertung: Mittel, da Stakeholder nicht regelmäßig informiert werden, kann dies zu Missverständnissen und zu einem Vertrauensverlust führen. Es könnte auch zu einer ineffizienten Nutzung des Systems kommen.

11.3. DARF: Die Transparenz darf an die jeweilige Zielgruppe angepasst werden, solange die Kernaussagen korrekt und nachvollziehbar bleiben.

Grundlegende Funktionalität: Die Transparenz sollte entsprechend der Zielgruppe angepasst werden, sodass alle relevanten Informationen klar, aber nicht überfordernd vermittelt werden.

Vorgesehener Einsatzkontext: Verwendung durch verschiedene Zielgruppen (z.B. technische Nutzer, Endkunden, Geschäftsleitung), die unterschiedliche Anforderungen an die Tiefe der Informationen haben.

Struktur der KI-Anwendung: Ein vielseitiges System, das von technisch versierten Nutzern und Laien gleichermaßen verwendet wird.

Maßnahmen:

Daten: Bereitstellung unterschiedlicher Detaillierungsgrade in der Dokumentation und Erklärung der Datennutzung, abgestimmt auf das Wissen der Zielgruppe und Erstellung von Datenberichten für unterschiedliche Zielgruppen (z.B. technische Berichte für Entwickler und Management-Zusammenfassungen für die Geschäftsleitung).

KI-Komponente: Anpassung der Erklärungsmechanismen je nach Zielgruppe: detaillierte technische Analysen für Experten, einfache Erklärungen oder Analogien für Laien und Entwicklung von mehrstufigen Erklärungen, die es erlauben, Informationen schrittweise zu vertiefen.

Einbettung: Bereitstellung von Benutzeroberflächen, die den Nutzer entsprechend seines Wissensstandes durch die Funktionalität der KI führen und anpassbare APIs und Dashboards für unterschiedliche Benutzerprofile.

Betrieb: Schulung und Aufklärung der Zielgruppen über die für sie relevanten Informationen. Regelmäßige Überprüfung, ob die bereitgestellten Informationen für die Zielgruppen verständlich und nützlich sind.

Risikobewertung: Niedrig, da die Anpassung der Transparenz an die Zielgruppe ist wichtig, aber ein Versäumnis führt selten zu gravierenden Folgen, solange die Kerninformationen korrekt vermittelt werden.

11.4. DARF NICHT: Wesentliche Informationen über die Funktionsweise des KI-Systems dürfen nicht zurückgehalten oder verschleiert werden, insbesondere wenn sie die Entscheidungen maßgeblich beeinflussen.

Grundlegende Funktionalität: Alle wesentlichen Informationen zur Funktionsweise der KI müssen offen offengelegt werden, um sicherzustellen, dass keine relevanten Fakten, die Entscheidungen beeinflussen, zurückgehalten werden.

Vorgesehener Einsatzkontext: Kritische Umgebungen, in denen Entscheidungen erhebliche Folgen haben (z.B. Justizsysteme, autonome Fahrzeuge, Gesundheitswesen).

Struktur der KI-Anwendung: Hochgradig komplexe KI-Systeme, die möglicherweise zu kritischen Entscheidungen führen, bei denen das vollständige Verständnis entscheidend ist.

Maßnahmen:

Daten: Vollständige Offenlegung der Herkunft der Daten, ihrer Vorverarbeitung und möglicher Schwachstellen oder Verzerrungen (z.B. Bias in Trainingsdaten) und Sicherstellung, dass keine für die Entscheidungsfindung relevanten Informationen verborgen werden.

KI-Komponente: Sicherstellung, dass alle wesentlichen Entscheidungsparameter offengelegt werden, selbst wenn dies die Komplexität des Modells offenbart sowie regelmäßige Validierung und Überprüfung des Systems, um sicherzustellen, dass keine relevanten Informationen verschleiert werden.

Einbettung: Benutzeroberflächen und Dokumentationen, die den Zugriff auf alle wesentlichen Informationen zur Funktionsweise des Systems ermöglichen.

Betrieb: Implementierung von Richtlinien, die es unmöglich machen, wesentliche Informationen absichtlich zu verbergen, sowie den regelmäßigen Audits, um sicherzustellen, dass Transparenz gewahrt wird.

Risikobewertung: Sehr hoch, da das Zurückhalten wesentlicher Informationen könnte zu schweren ethischen, rechtlichen oder sicherheitsrelevanten Problemen führen. Und es könnten falsche Entscheidungen getroffen werden, die auf unvollständigen oder falschen Informationen basieren.

12. Autonomie (Autonomy):

12.1. MUSS: KI-Systeme müssen so gestaltet werden, dass sie ihre Aufgaben autonom ausführen können, aber stets innerhalb der vom Menschen festgelegten Grenzen und unter menschlicher Aufsicht.

Grundlegende Funktionalität: Autonome Ausführung von Aufgaben mit menschlicher Aufsicht.

Vorgesehener Einsatzkontext: Anwendungen, bei denen kontinuierliche menschliche Überwachung möglich ist, z. B. industrielle Automatisierung oder medizinische Diagnosesysteme.

Struktur der KI-Anwendung: Ein überwachtes lernendes System mit klaren Grenzbedingungen, die durch menschliche Eingriffe verstärkt werden können.

Maßnahmen:

Daten: Sicherstellen, dass die Trainingsdaten repräsentativ sind und die relevanten Grenzfälle abdecken. Implementierung von Mechanismen, die Warnungen bei Ausreißern oder unerwarteten Ergebnissen geben.

KI-Komponente: Die KI muss so programmiert sein, dass sie stets innerhalb vordefinierter Parameter arbeitet. Es sollten Mechanismen vorhanden sein, die eine Rückkopplung zur menschlichen Aufsicht sicherstellen.

Einbettung: Implementierung von Kontrollmechanismen, die es dem Menschen ermöglichen, die KI jederzeit zu unterbrechen oder die Parameter anzupassen.

Betrieb: Regelmäßige Überprüfung und Anpassung der KI-Parameter durch menschliche Operatoren. Integration von Protokollen, die Berichte über Entscheidungen und Handlungen der KI-Systeme bereitstellen.

Risikobewertung: Mittel, da die menschliche Aufsicht minimiert das Risiko, aber es besteht das Risiko von Überlastung oder Unaufmerksamkeit seitens des menschlichen Operators.

12.2. SOLLTE: Die Autonomie von KI-Systemen sollte dem jeweiligen Anwendungsfall angemessen sein, sodass sie effizient agieren, ohne menschliches Eingreifen in Routineaufgaben zu erfordern.

Grundlegende Funktionalität: Effiziente Autonomie für Routineaufgaben ohne menschliches Eingreifen.

Vorgesehener Einsatzkontext: Automatisierung von Routineprozessen in Bereichen wie Logistik, einfache Entscheidungsunterstützung, IT-Systeme.

Struktur der KI-Anwendung: Systeme mit niedriger Komplexität und klaren Aufgaben, z. B. Regel-basierte Systeme oder einfache Machine-Learning-Algorithmen.

Maßnahmen:

Daten: Regelmäßige Aktualisierung der Trainingsdaten, um sicherzustellen, dass Routineaufgaben weiterhin ohne Fehler durchgeführt werden.

KI-Komponente: Die KI sollte klar zwischen Routineaufgaben und Sonderfällen unterscheiden können. Integration von Mechanismen zur Erkennung von Anomalien.

Einbettung: Es sollte ein klarer Übergang zwischen autonomem Handeln und menschlichem Eingreifen möglich sein, insbesondere bei nicht-routinemäßigen Aufgaben.

Betrieb: Sicherstellen, dass die Routineaufgaben weiterhin angemessen funktionieren, indem kontinuierlich Performance-Überprüfungen durchgeführt werden.

Risikobewertung: Niedrig, da nur Routineaufgaben autonom ausgeführt werden, besteht ein geringes Risiko schwerwiegender Fehler, sodass das Risiko steigt, wenn Sonderfälle nicht richtig erkannt werden.

12.3. DARF: Ein KI-System darf in bestimmten festgelegten Bereichen Entscheidungen ohne menschliches Zutun treffen, wenn diese Bereiche klar definiert und überwacht sind.

Grundlegende Funktionalität: Autonome Entscheidungen in vorab definierten Bereichen, die überwacht werden.

Vorgesehener Einsatzkontext: Systeme, die in eng umgrenzten Bereichen agieren, z. B. autonome Fahrzeuge in klar markierten Gebieten oder Finanzsysteme mit festgelegten Regeln.

Struktur der KI-Anwendung: Komplexere Algorithmen, die in spezialisierten Umgebungen agieren, z. B. Reinforcement Learning oder Deep Learning mit klar definierten Constraints.

Maßnahmen:

Daten: Regelmäßige Validierung der Datensätze, um sicherzustellen, dass sie den festgelegten Entscheidungsbereichen entsprechen. Überwachung der Datenqualität und Auffälligkeiten.

KI-Komponente: Die KI muss auf Abweichungen aufmerksam machen und im Falle von Unsicherheiten oder Grenzfällen eine menschliche Eingriffsmöglichkeit bieten.

Einbettung: Die Überwachung der definierten Entscheidungsbereiche muss lückenlos sein, damit bei Abweichungen menschliche Eingriffe sofort möglich sind.

Betrieb: Implementierung von Mechanismen, die es ermöglichen, Entscheidungen im Nachhinein zu bewerten und eventuell anzupassen. Überwachungsprotokolle sollten kontinuierlich geführt werden.

Risikobewertung: Mittel bis Hoch, da es stark davon abhängt, wie gut die Entscheidungsbereiche definiert sind und wie effektiv die Überwachungsmechanismen arbeiten. Mangelnde Überwachung kann zu unerwünschten autonomen Handlungen führen.

12.4. DARF NICHT: KI-Systeme dürfen keine vollständig autonomen Entscheidungen in sicherheitskritischen oder ethisch sensiblen Bereichen treffen, ohne dass eine menschliche Überprüfung vorgesehen ist.

Grundlegende Funktionalität: KI-Systeme dürfen in sicherheitskritischen oder ethischen Bereichen nicht autonom agieren.

Vorgesehener Einsatzkontext: Gesundheitswesen, Justiz, Militär, Luftfahrt und andere sicherheitskritische Bereiche.

Struktur der KI-Anwendung: Hochkomplexe Systeme, die potenziell sicherheitskritische oder ethische Implikationen haben könnten, wie z. B. prädiktive Modelle oder autonome Waffensysteme.

Maßnahmen:

Daten: Strenge Kontrolle und Validierung der Datenquellen, um sicherzustellen, dass sie keine ethischen Bedenken aufwerfen und die Sicherheit nicht gefährden.

KI-Komponente: Implementierung von Mechanismen, die sicherstellen, dass keine autonomen Entscheidungen getroffen werden, sondern dass stets eine menschliche Überprüfung und Freigabe erforderlich ist.

Einbettung: In sicherheitskritischen Bereichen müssen Notfallprotokolle vorhanden sein, die es ermöglichen, die KI zu stoppen oder zu überschreiben, wenn es zu kritischen Situationen kommt.

Betrieb: Regelmäßige Überprüfung der Algorithmen durch Menschen und Audits der sicherheitsrelevanten Entscheidungen und Menschen sollten immer in den Entscheidungsprozess involviert bleiben.

Risikobewertung: Sehr hoch, da ein Versagen der menschlichen Überprüfung oder der Kontrollmechanismen in sicherheitskritischen oder ethisch sensiblen Bereichen katastrophale Folgen haben.

13. Kontrolle (Control):

13.1. MUSS: Menschen müssen jederzeit die Möglichkeit haben, die Kontrolle über ein KI-System zu übernehmen, insbesondere in Situationen, in denen das Verhalten des Systems unvorhersehbar oder risikoreich ist.

Grundlegende Funktionalität: Das KI-System muss eine Möglichkeit bieten, manuell eingreifen zu können, besonders bei unvorhersehbaren oder risikoreichen Situationen.

Vorgesehener Einsatzkontext: Systeme, die in sicherheitskritischen Bereichen wie Medizin, autonome Fahrzeuge oder Industrie eingesetzt werden.

Struktur der KI-Anwendung: Die Architektur muss einen „Override“-Modus beinhalten, der es Menschen erlaubt, in das System einzugreifen und es zu steuern, wenn es risikoreich agiert sowie dies könnte durch eine spezielle Steuerungsschicht innerhalb der Software ermöglicht werden, die sicherstellt, dass die menschliche Kontrolle immer Vorrang hat.

Maßnahmen:

Daten: Implementierung von Daten-Logging-Mechanismen zur Nachverfolgung von Entscheidungen und zur Identifikation von Anomalien.

KI-Komponente: Überwachung der Entscheidungsprozesse mit einem Fokus auf Fehlverhalten oder unerwartete Ergebnisse.

Einbettung: Schnittstellen zur sofortigen Übernahme durch Menschen, z. B. ein „Notaus“-Schalter oder eine Pausierungsmöglichkeit.

Betrieb: Regelmäßige Überprüfung des Systems auf die Fähigkeit, vom Menschen übernommen zu werden, besonders in risikoreichen Szenarien.

Risikobewertung: **Hoch**, da in sicherheitskritischen Bereichen können Fehlfunktionen katastrophale Folgen haben.

13.2. SOLLTE: Es sollte eine klare und intuitive Benutzerschnittstelle vorhanden sein, die es dem Menschen erlaubt, das KI-System einfach zu überwachen, anzupassen und bei Bedarf zu stoppen. Bei Änderungen innerhalb der KI-Anwendungen und Nutzer sollten Risiken neu betrachtet werden.

Grundlegende Funktionalität: Das System sollte einfach verständliche Schnittstellen bereitstellen, die kontinuierliche Überwachung und manuelle Anpassungen ermöglichen.

Vorgesehener Einsatzkontext: Systeme mit langfristiger Überwachung wie in der Industrie oder in Verkehrsleitsystemen.

Struktur der KI-Anwendung: Die Architektur sollte modular aufgebaut sein, um Änderungen oder Anpassungen an der Schnittstelle zu ermöglichen sowie dies ermöglicht es, die Benutzeroberfläche flexibel an die Anforderungen der Nutzer anzupassen, ohne die Kernkomponenten der KI zu gefährden.

Maßnahmen:

Daten: Visualisierung der Systemaktivitäten, wie z. B. der Entscheidungsflüsse und Auffälligkeiten in den Daten.

KI-Komponente: Adaptive Mechanismen, die es dem Nutzer ermöglichen, Parameter und Prozesse der KI flexibel anzupassen.

Einbettung: Eine intuitive Benutzeroberfläche, die den Zustand des Systems in Echtzeit anzeigt und klare Eingriffsmöglichkeiten bietet.

Betrieb: Regelmäßige Schulungen der Nutzer, um sicherzustellen, dass sie in der Lage sind, Änderungen vorzunehmen und Risiken richtig einzuschätzen.

Risikobewertung: **Mittel**, da ein mangelhaft gestaltetes Interface kann zu Fehlbedienungen führen, was je nach Anwendungsgebiet unterschiedlich kritisch ist.

13.3. DARF: Kontrollmechanismen dürfen automatisierte Notabschaltungen oder Fail-Safe-Protokolle enthalten, um Schäden oder Fehlfunktionen zu vermeiden.

Grundlegende Funktionalität: Das System sollte in der Lage sein, automatisch abzuschalten oder in einen sicheren Zustand zu versetzen, wenn eine Fehlfunktion festgestellt wird.

Vorgesehener Einsatzkontext: Autonome Systeme in der Industrie, dem Verkehr und der Medizin.

Struktur der KI-Anwendung: Es sollten redundante Steuerungsmechanismen eingebaut werden, die sicherstellen, dass das System bei einem Fehler kontrolliert heruntergefahren wird und dazu gehören Backups, Watchdog-Systeme und mechanische Fail-Safe-Protokolle.

Maßnahmen:

Daten: Erhebung von Umgebungs- und Systemdaten, um Fehlverhalten frühzeitig zu erkennen.

KI-Komponente: Entwicklung von Fail-Safe-Protokollen, die auf erkannte Risiken reagieren und das System in einen sicheren Zustand versetzen.

Einbettung: Integration automatisierter Abschaltmechanismen und Sicherheitsmodi in das Gesamtsystem.

Betrieb: Implementierung regelmäßiger Tests der Fail-Safe-Protokolle, um sicherzustellen, dass sie im Ernstfall korrekt funktionieren.

Risikobewertung: **Mittel**, da automatisierte Abschaltungen könnten in kritischen Momenten ausgelöst werden, jedoch minimieren sie das Risiko eines größeren Schadens.

13.4. DARF NICHT: Ein KI-System darf nicht so gestaltet sein, dass es der menschlichen Kontrolle entzogen oder sich unkontrollierbar weiterentwickelt, ohne dass die menschlichen Akteure eingreifen können.

Grundlegende Funktionalität: Das System muss so gestaltet sein, dass es sich nicht außerhalb des vorgesehenen Bereichs ohne menschliches Eingreifen weiterentwickeln kann.

Vorgesehener Einsatzkontext: Systeme in allen sicherheitsrelevanten oder regulierten Bereichen, z. B. in der Medizin oder im Militär.

Struktur der KI-Anwendung: Die KI-Struktur sollte beschränkt und kontrolliert entwickelt werden, sodass selbstlernende Mechanismen nur in definierten Bereichen und unter menschlicher Aufsicht agieren können sowie diese schließt die Implementierung von Grenzwerten und Prüfmechanismen ein.

Maßnahmen:

Daten: Implementierung strikter Datengovernance-Richtlinien, um zu verhindern, dass das System Daten verwendet, die zu unerwarteten Ergebnissen führen könnten.

KI-Komponente: Begrenzung der Lernfähigkeit der KI, um selbstständige Evolution oder Modifikation ohne menschliche Zustimmung zu verhindern.

Einbettung: Manuelle Kontrollpunkte, die einen menschlichen Eingriff bei wesentlichen Änderungen erfordern.

Betrieb: Einrichtung eines Überwachungssystems, das sicherstellt, dass sich das System nicht unvorhergesehen verändert oder weiterentwickelt.

Risikobewertung: Sehr hoch, da ein unkontrollierbares System könnte unvorhersehbare und potenziell katastrophale Folgen haben.

Hinweis: Diese Regeln bieten einen Rahmen, der sicherstellt, dass KI-Systeme verantwortungsvoll, sicher, fair und transparent entwickelt und eingesetzt werden sowie weitere Regeln können ergänzend erstellt werden!

Anhang II Die Beschreibung der Bausteine - KI-Lifecycle-Process.

Beschreibung der Bausteine - KI-Lifecycle-Process wurden mit der subjektiven Betrachtung anlehnend gemäss CellPress open Access¹⁹⁷ in zusammengefasster Form erstellt: (Bewertung wurde erstellt: grundlegende Funktionalität, vorgesehener Einsatzkontext, der Struktur der KI-Anwendung mit erstelle Massnahmen für Daten, KI-Komponente, Einbettung und für den Betrieb sowie Risikobewertung (niedrig, mittel, hoch, sehr hoch)).

Design AI/Data Scientist

1. Problem identifizieren und formuliert werden: Das zugrunde liegende Problem MUSS klar und präzise formulieren, sodass die nachfolgenden Phasen der Modellierung und Integration effektiv erstellt werden können. Dabei SOLLTEN die richtigen Modelle und Algorithmen ausgewählt werden und es DARF dabei auch eine flexible Anpassung der Problemstellung je nach Entwicklung neue Erkenntnisse vorgenommen werden.

Grundlegende Funktionalität: Das zugrunde liegende Problem, das durch die KI gelöst werden soll, identifiziert und klar formuliert mit dem Ziel ist es sicherzustellen, dass das Problem präzise und verständlich beschrieben wird, damit später die passende Modellierungsstrategie und Algorithmen ausgewählt werden können.

Vorgesehener Einsatzkontext: Der Einsatzkontext bestimmt, welche Art von Modell oder Lösung entwickelt werden soll – beispielsweise ob es um die Klassifizierung von Objekten, die Vorhersage von Zeitreihen oder die Segmentierung von Kundengruppen geht und diese Phase legt auch fest, welche Ergebnisse von der KI erwartet werden und in welchem Umfeld sie eingesetzt wird (z.B. im industriellen, medizinischen oder sozialen Bereich).

Struktur der KI-Anwendung: Die Struktur der KI-Anwendung hängt von der Natur des identifizierten Problems ab und es kann sich um einfache Modelle (wie lineare Regression) oder komplexere Modelle (wie neuronale Netze) handeln, je nach Anwendungsfall und den zur Verfügung stehenden Daten.

Maßnahmen:

Daten: Sicherstellen, dass die zur Verfügung stehenden Daten ausreichen, um das Problem zu adressieren und dazu gehört die Überprüfung, ob die Daten relevant und umfassend genug sind, um das Problem zu lösen.

KI-Komponente: Auswahl geeigneter Algorithmen und Methoden, die dem Problem angemessen sind sowie die Komplexität des Problems bestimmt die Wahl der Modellierungsansätze.

Einbettung: Bewertung, ob die formulierte Lösung in bestehende Systeme oder Arbeitsprozesse integriert werden kann und die Kompatibilität und Interoperabilität müssen sichergestellt sein.

Betrieb: Sicherstellen, dass die Lösung skalierbar und belastbar genug ist, um im geplanten Betrieb reibungslos zu funktionieren sowie dabei müssen auch Fragen der Wartbarkeit und der technischen Betreuung geklärt werden.

Risikobewertung: Mittel, da das Problem entweder zu komplex oder zu simpel zu formulieren, was zu ineffektiven Lösungen führen könnte und da die genaue Formulierung des Problems entscheidend für den weiteren Entwicklungsverlauf ist, kann ein Fehler in dieser Phase später zu hohen Kosten führen sowie dennoch ist es in dieser Phase noch relativ einfach, Korrekturen vorzunehmen.

2. Daten- und KI-Ethik überprüfen: Die Umsetzung MUSS zwingend auf ethische Aspekte bei der Datenverwendung und bei der Modellentwicklung auf Fairness, Datenschutz und Nichtdiskriminierung geachtet werden. Dabei SOLLTE eine kontinuierliche Überprüfung dieser ethischen Standards im Betrieb erfolgen sowie DARF auch zusätzliche ethische Prüfungen in bestimmte Anwendungsbereiche erstellt

vgl.¹⁹⁷ (Access)

werden und dabei DARF es NICHT zu rechtlichen Konsequenzen und Vertrauensverluste führen.

Grundlegende Funktionalität: Die Überprüfung der ethischen Fragen, die sich bei der Verwendung von Daten und bei der Entwicklung der KI ergeben und soll sichergestellt werden, dass die Nutzung der Daten und die Ergebnisse der KI mit ethischen Prinzipien und gesetzlichen Vorschriften übereinstimmen, insbesondere im Hinblick auf Datenschutz, Fairness, Transparenz und Nichtdiskriminierung.

Vorgesehener Einsatzkontext: Der Einsatzkontext bestimmt, welche ethischen Fragen besonders wichtig sind und so müssen beispielsweise im medizinischen Kontext strenge Vorschriften zum Schutz der Patientendaten eingehalten werden, während in sozialen Medien der Schutz der Privatsphäre und die Vermeidung von Diskriminierung im Vordergrund stehen.

Struktur der KI-Anwendung: Die ethischen Überlegungen beeinflussen die Auswahl der Algorithmen und die Struktur der KI, z.B. müssen Algorithmen so gestaltet sein, dass sie keine Verzerrungen (Bias) gegenüber bestimmten Gruppen aufweisen und die Entscheidungen nachvollziehbar bleiben.

Maßnahmen:

Daten: Überprüfen, ob die gesammelten Daten fair und repräsentativ sind und es bedeutet, dass die Daten keine Voreingenommenheit enthalten und alle relevanten Gruppen angemessen repräsentieren.

KI-Komponente: Sicherstellen, dass die verwendeten KI-Algorithmen keine systematischen Verzerrungen aufweisen, die bestimmte Gruppen benachteiligen könnten und es sollte auch geprüft werden, ob die Algorithmen transparent und nachvollziehbar sind.

Einbettung: Einführung von Mechanismen zur kontinuierlichen Überprüfung, ob die KI im Einsatz ethische Standards einhält sowie Auditing- und Überwachungssysteme sind notwendig, um sicherzustellen, dass die Anwendung im Laufe der Zeit keine unethischen Ergebnisse liefert.

Betrieb: Implementierung von Prozessen, die es ermöglichen, ethische Bedenken im laufenden Betrieb zu identifizieren und zu beheben und dazu gehören Mechanismen für Benutzerrückmeldungen, ethische Schulungen und regelmäßige Audits.

Risikobewertung: Hoch, da die Missachtung ethischer Prinzipien schwerwiegende rechtliche und gesellschaftliche Konsequenzen haben kann sowie ein KI-System, das etwa diskriminierende Entscheidungen trifft, kann zu erheblichen Imageschäden und rechtlichen Auseinandersetzungen führen.

3. Technisches Wissen zu KI-Algorithmen, Anwendungen und relevanten Modellen überprüfen: Technische Neuerungen SOLLTEN auf dem aktuellen Stand der Technik von modernen und bewährten Algorithmen und Methoden oder Modellen eingesetzt werden.

Grundlegende Funktionalität: Die Überprüfung der technischen Literatur stellt sicher, dass aktuelle Methoden und Best Practices berücksichtigt werden sowie umfasst das Studium von wissenschaftlichen Veröffentlichungen, technischen Berichten und Dokumentationen zu KI-Algorithmen, um sicherzustellen, dass die Wahl der Algorithmen und der Modellarchitekturen den neuesten Stand der Technik widerspiegelt.

Vorgesehener Einsatzkontext: Im vorgesehenen Einsatzkontext sollen die bestmöglichen Methoden und Modelle angewendet werden, um eine robuste und effiziente Lösung zu entwickeln. Die Anwendung sollte der spezifischen Domäne entsprechen und aktuelle Forschungsergebnisse integrieren.

Struktur der KI-Anwendung: Die Struktur der KI-Anwendung wird durch die in der Literatur gefundenen Informationen beeinflusst und dies könnte die Wahl zwischen verschiedenen Modelltypen (z.B. Entscheidungsbäume, neuronale Netze, Support-Vektor-Maschinen) sowie die Feinabstimmung der Algorithmen umfassen.

Maßnahmen:

Daten: Sicherstellen, dass die Algorithmen mit den vorhandenen Daten arbeiten können. Falls neue Modelle oder Algorithmen verwendet werden, sollten sie mit den spezifischen Datenarten (z.B. Text, Bild, Zeitreihen) kompatibel sein.

KI-Komponente: Auswahl von Algorithmen und Modellen, die nicht nur den aktuellen Stand der Technik widerspiegeln, sondern auch effizient und robust genug sind, um die gestellten Anforderungen zu erfüllen.

Einbettung: Sicherstellen, dass die Modelle und Algorithmen leicht in die bestehenden Systeme integriert werden können. Hierbei sollte auch auf die technische Kompatibilität geachtet werden.

Betrieb: Sicherstellen, dass die Implementierung aktueller Forschungsergebnisse und neuer Algorithmen im laufenden Betrieb überwacht wird und dazu gehört auch die regelmäßige Überprüfung, ob neue Forschungsergebnisse oder Updates der Algorithmen verfügbar sind.

Risikobewertung: Mittel, da fehlerhafte oder veraltete Wahl der Algorithmen kann die Leistung und Effizienz der KI-Anwendung beeinträchtigen, jedoch ist dies in der Regel frühzeitig erkennbar und korrigierbar.

4. Datenaufbereitung: Die Daten MÜSSEN korrekt bereinigt, transformiert und normalisiert werden, um eine hervorragende Basis für das Modelltrainingsqualität und -richtigkeit zu erstellen und dabei SOLLTE eine automatisierte Datenaufbereitungs-pipelines in Prozess integriert sein, um den Prozess effizient zu gestalten und um eine bestmögliche Modelleistung zu bekommen. Dabei DARF die Verwendung von unterschiedlichen Methoden der Datenaufbereitung getestet werden, um die bestmögliche Modelleistung zu erzielen. Hierbei DÜRFEN NICHT Fehler übersehen werden, da diese zu unkorrekten oder fehlerhaften Modellen führen können.

Grundlegende Funktionalität: Datenaufbereitung ist ein entscheidender Schritt im KI-Entwicklungsprozess und umfasst das Bereinigen, Normalisieren, Transformieren und Formatieren der Daten, sodass sie für das Modell trainingsbereit sind und hierbei werden auch fehlende Daten behandelt und potenzielle Fehler oder Ausreißer erkannt und korrigiert.

Vorgesehener Einsatzkontext: Im vorgesehenen Einsatzkontext müssen die Daten so aufbereitet werden, dass sie für das spezifische Modell und die spezifische Anwendung geeignet sind mit dem Fokus liegt darauf, qualitativ hochwertige, konsistente und fehlerfreie Daten zur Verfügung zu stellen.

Struktur der KI-Anwendung: Die Datenaufbereitung beeinflusst die Struktur der KI-Anwendung erheblich. mit falsch aufbereitete Daten können zu ungenauen Modellen und Fehlvorhersagen führen sowie eine korrekte Aufbereitung ist essenziell für die Effektivität und Genauigkeit des KI-Systems.

Maßnahmen:

Daten: Implementierung robuster Datenaufbereitungs-pipelines, die automatisiert Daten bereinigen, transformieren und validieren sowie fehlende Werte, Ausreißer und inkonsistente Daten sollten korrekt behandelt werden.

KI-Komponente: Sicherstellen, dass das KI-Modell auf den aufbereiteten Daten gut performt und unterschiedliche Aufbereitungsmethoden (z.B. Normalisierung, Skalierung) sollten an das spezifische Modell angepasst werden.

Einbettung: Die Datenaufbereitung sollte in die bestehenden Datensysteme und Pipeline-prozesse eingebettet werden, um eine nahtlose Integration zu gewährleisten.

Betrieb: Überwachung der Datenaufbereitung im laufenden Betrieb und es sollte sichergestellt werden, dass neu eingehende Daten korrekt und konsistent verarbeitet werden.

Risikobewertung: Hoch, da Fehler in der Datenaufbereitung führen direkt zu fehlerhaften Modellen und falschen Ergebnissen sowie eine mangelhafte Datenqualität beeinträchtigt die Modellgenauigkeit erheblich, weshalb in diesem Schritt ein hohes Risiko besteht.

5. Datenexploration: Es SOLLTEN die Struktur und die Eigenschaften sowie die Muster verständlich sein, um alle Anomalien frühzeitig zu erkennen und zu bereinigen.

Grundlegende Funktionalität: Die Datenexploration umfasst die erste Analyse der verfügbaren Daten, um Muster, Zusammenhänge und Anomalien zu identifizieren und dies hilft, ein besseres Verständnis der Daten zu erlangen und herauszufinden, welche Merkmale für das Modell wichtig sein könnten.

Vorgesehener Einsatzkontext: Im vorgesehenen Einsatzkontext dient die Datenexploration dazu, Hypothesen zu testen und die Daten besser zu verstehen, bevor sie in Modelle einge-

speist werden und dies ist ein wichtiger Schritt, um sicherzustellen, dass das Modell auf einer soliden Datenbasis aufbaut.

Struktur der KI-Anwendung: Die Ergebnisse der Datenexploration beeinflussen die spätere Modellauswahl und die Feinabstimmung der Modelle sowie durch das Erkennen von Korrelationen und Mustern können spezifische Modellparameter optimiert werden.

Maßnahmen:

Daten: Verwenden von statistischen Methoden und Visualisierungstools, um die Datenstruktur zu analysieren sowie Fehler, Ausreißer oder unerwartete Muster sollten frühzeitig identifiziert werden.

KI-Komponente: Sicherstellen, dass die explorativen Analysen genutzt werden, um das Modell-Feature-Engineering zu verbessern sowie Datenexploration kann dazu beitragen, die besten Merkmale für das Modell auszuwählen.

Einbettung: Die Erkenntnisse aus der Exploration sollten in den Entwicklungsprozess eingebettet werden und dies kann durch Anpassungen in den Datenpipelines oder durch Änderungen an der Modellarchitektur erfolgen.

Betrieb: Implementierung automatisierter Mechanismen zur kontinuierlichen Datenüberwachung und Exploration, um sicherzustellen, dass neu hinzukommende Daten keine unerwarteten Änderungen im Modell verursachen.

Risikobewertung: Mittel, da Fehler in der Exploration in späteren Phasen der Modellierung schwerer zu erkennen sind sowie eine fehlerhafte Exploration kann dazu führen, dass wichtige Muster übersehen werden oder irrelevante Daten in das Modell einfließen.

6. Externe Datenexploration: Es MUSS die Qualität und Kompatibilität von externen Daten (auch interne Daten) sorgfältig untersucht werden, bevor diese in das Modell implementiert und diese Dateien SOLLTEN kontinuierlich überwacht werden, um die Modelleistung und die Genauigkeit zu erhöhen.

Grundlegende Funktionalität: Es werden externe Datenquellen untersucht, die die internen Daten ergänzen können und dies ist besonders wichtig, wenn interne Daten unvollständig oder zu spezifisch sind, um robuste Vorhersagen zu treffen.

Vorgesehener Einsatzkontext: Die Nutzung externer Daten kann die Modelleistung erheblich verbessern, indem zusätzliche Informationen bereitgestellt werden und dies ist besonders in Kontexten wichtig, in denen vielfältige oder alternative Datenquellen verfügbar sind, wie z.B. öffentliche Datensätze oder Drittanbieter-Daten.

Struktur der KI-Anwendung: Die Struktur der KI-Anwendung muss so angepasst werden, dass sie externe Datenquellen effektiv integriert und dies erfordert möglicherweise die Anpassung der Datenaufbereitungspipelines sowie eine Validierung der Qualität und Konsistenz externer Daten.

Maßnahmen:

Daten: Bewertung der Qualität und Vertrauenswürdigkeit externer Daten. Und es muss sichergestellt werden, dass externe Daten kompatibel und frei von Verzerrungen oder ethischen Bedenken sind.

KI-Komponente: Sicherstellen, dass die Algorithmen korrekt mit den externen Daten arbeiten und dass die Integration keine negativen Auswirkungen auf die Modelleistung hat.

Einbettung: Externe Datenquellen sollten in bestehende Pipelines integriert werden, ohne den Datenfluss oder die bestehenden Prozesse zu stören sowie Datenintegrationsstrategien sind hier besonders wichtig.

Betrieb: Überwachung der externen Datenquellen im laufenden Betrieb sowie externe Datenquellen können sich ändern oder sogar unzuverlässig werden, was kontinuierlich überwacht werden muss.

Risikobewertung: Sehr Hoch, da die Verwendung externer Datenquellen birgt ein hohes Risiko und diese Daten sind möglicherweise nicht konsistent oder von minderer Qualität, was zu ungenauen Vorhersagen oder Verzerrungen führen kann, wobei könnten rechtliche und ethische Fragen bezüglich der Nutzung externer Daten auftreten.

Develop AI/ML Scientist

7. Datenexploration: Die Datenstruktur MUSS analysiert werden, um Ausreißer in den Mustern, Anomalien oder fehlende Werte zu erkennen und somit eine solide Modellierungsgrundlage zu erstellen. Hierbei helfen Visualisierungstools und statistische Methoden verwenden, um das Datenverständnis zu verbessern und frühzeitig Mustererkennung schneller zu entdecken. Datenprobleme DARF NICHT ignoriert oder nur oberflächlich analysiert werden, da diese im späteren Prozess zu Verzerrungen und schlechten Modellergebnissen führen.

Grundlegende Funktionalität: Die Datenexploration ist der erste Schritt im Prozess der Datenvorbereitung, bei dem die verfügbaren Daten gründlich analysiert werden und hierbei geht es darum, ein Verständnis für die Struktur und den Inhalt der Daten zu erlangen und dies umfasst die Überprüfung von Datentypen, das Auffinden von Ausreißern, das Erkennen von fehlenden Werten und die Analyse der Verteilungen sowie das Ziel ist es, mögliche Probleme oder Unregelmäßigkeiten in den Daten frühzeitig zu identifizieren, um später bessere Modelle zu entwickeln.

Vorgesehener Einsatzkontext: Die Datenexploration wird in der Vorbereitungsphase eines Projekts durchgeführt und hat großen Einfluss auf die nachfolgenden Schritte und dies hilft dabei, Probleme in den Daten zu erkennen, die während der Modellerstellung zu Verzerrungen führen könnten sowie das Ziel ist es, ein tiefes Verständnis der Daten zu erlangen, bevor komplexe Modelle erstellt werden.

Struktur der KI-Anwendung: In dieser Phase wird kein spezifisches KI-Modell entwickelt und stattdessen kommen Visualisierungstools (wie Matplotlib oder Seaborn), statistische Analysen (z. B. Mittelwert, Median, Varianz) und Explorative Data Analysis (EDA) zum Einsatz sowie werden Algorithmen für die Clusterbildung, Korrelationen oder statistische Tests genutzt, um die Daten besser zu verstehen.

Maßnahmen:

Daten: Bereinigung der Daten, Umgang mit fehlenden Werten und Identifikation von Anomalien oder Ausreißern.

KI-Komponente: Kein spezifisches KI-Modell erforderlich, aber erste Analysen zur Feststellung von Korrelationen und Mustern können durchgeführt werden.

Einbettung: Verwendung von Analysetools wie Jupyter Notebooks, R oder Python für die Exploration der Daten.

Betrieb: Sicherstellen, dass regelmäßige Überprüfungen der Datenqualität durchgeführt werden, besonders bei der Erfassung neuer Daten.

Risikobewertung: Mittel, da unentdeckte Probleme in den Daten später zu Fehlinterpretationen oder Verzerrungen führen können und eine sorgfältige Datenanalyse ist notwendig, um das Risiko späterer Modellfehler zu minimieren.

8. Erstes KI-Modell erstellen: Eine Erstellung eines einfachen Modells MUSS auf Machbarkeit mit den vorhandenen Daten getestet werden, um vorliegende Muster zu erkennen. Dabei SOLLTE eine überschaubare Datenmenge trainiert und getestet werden im Modell, um erste Kenntnisse über die Modellleistung zu bekommen. Hierbei DARF die Verwendung unterschiedliche Algorithmen eingesetzt werden, um die perfekte geeignete Methode zu erhalten bis der Prototyp zum endgültigen Lösungsweg führt.

Grundlegende Funktionalität: Das Ziel ist es, ein erstes, einfaches KI-Modell zu erstellen und dient oft als Proof-of-Concept, um festzustellen, ob die gewählten Daten und Algorithmen geeignet sind, das Problem zu lösen und einfache Modelle wie Lineare Regression, Entscheidungsbäume oder k-Nächste-Nachbarn werden häufig verwendet, um die grundlegenden Muster in den Daten zu verstehen und eine erste Einschätzung der Machbarkeit zu geben.

Vorgesehener Einsatzkontext: Der Einsatzkontext ist die Entwicklung eines ersten Prototyps, wobei es hilft dabei, die Anforderungen und den Umfang des Problems besser zu verstehen und das Modell wird getestet, um zu sehen, ob es brauchbare Ergebnisse liefert, bevor komplexere Modelle oder Techniken in Betracht gezogen werden.

Struktur der KI-Anwendung: Das Modell wird auf einer kleinen, überschaubaren Datenmenge trainiert, um die grundlegenden Beziehungen zwischen den Variablen zu identi-

zieren und es handelt sich typischerweise um ein einfaches Modell, das leicht zu interpretieren ist und nur grundlegende Datenvorverarbeitung erfordert.

Maßnahmen:

Daten: Die Daten müssen vorverarbeitet werden, indem sie z. B. skaliert, normalisiert oder kategorisiert werden, um sicherzustellen, dass das Modell gut mit ihnen umgehen kann.

KI-Komponente: Ein einfaches, leicht interpretierbares Modell wird erstellt, um die Machbarkeit zu testen und es ist nicht erforderlich, ein komplexes Modell zu nutzen, solange grundlegende Muster erkennbar sind.

Einbettung: Diese Phase findet oft in einer Prototyping-Umgebung statt, wie z. B. Jupyter Notebooks oder einer Entwicklungsumgebung wie Spyder oder R-Studio.

Betrieb: Das Modell wird noch nicht produktiv eingesetzt und es bleibt auf einer Testumgebung beschränkt und dient als erster Evaluierungsschritt.

Risikobewertung: **Niedrig**, da es sich um einen Prototyp handelt, ist das Risiko gering und die Ergebnisse dieses Modells sind jedoch begrenzt, und die Erkenntnisse müssen mit Vorsicht interpretiert werden.

9. Datenaugmentation: Die Qualität von augmentierte Daten MUSS im Modell so verbessert werden, dass das Risiko von Overfitting minimiert wird und man SOLLTE Datenaugmentationstechniken einsetzt, wenn die Datenmenge begrenzt oder unausgewogen ist, um die Trainingsbasis zu erweitern. Dabei DÜRFEN unterschiedliche Augmentationstechniken getestet werden, um die perfekte quantitative Methode zur Ausweitung der Daten zu finden.

Grundlegende Funktionalität: Die Datenaugmentation ist eine Technik, die dazu dient, die verfügbare Datenmenge künstlich zu vergrößern und dies ist besonders hilfreich, wenn der Datensatz zu klein oder unausgewogen ist sowie typische Methoden umfassen das Hinzufügen von Rauschen, geometrische Transformationen (bei Bildern) oder synthetische Generierung von Datenpunkten (z. B. SMOTE für Klassifikationsaufgaben) und somit das Ziel, die Robustheit des Modells zu erhöhen und das Risiko des Overfittings zu verringern.

Vorgesehener Einsatzkontext: Diese Technik wird häufig in Bereichen eingesetzt, in denen Daten teuer oder schwer zu beschaffen sind, z. B. in der Bildverarbeitung, Textanalyse oder in der Medizin und bei unausgewogenen Datensätzen hilft die Augmentation, die Minderheitsklassen zu stärken.

Struktur der KI-Anwendung: Die Augmentation wird in der Vorverarbeitungspipeline implementiert und verändert die Trainingsdaten so, dass das Modell mit einer breiteren Palette von Beispielen trainiert wird und kann auf verschiedene Datentypen (Bilder, Texte, Zeitreihen) angewendet werden, um die Datenbasis zu erweitern.

Maßnahmen:

Daten: Sicherstellen, dass die Augmentation sinnvoll ist und keine unnatürlichen oder verzerrten Daten erzeugt und die augmentierten Daten müssen die Realität widerspiegeln, um das Modell nicht zu verwirren.

KI-Komponente: Das Modell muss in der Lage sein, mit den augmentierten Daten zu trainieren, ohne dass es zu Overfitting kommt sowie regelmäßige Validierungsschritte sind notwendig.

Einbettung: Integration der Augmentationstechniken in die Datenpipeline. Automatisierte Augmentation ist oft in Frameworks wie TensorFlow oder PyTorch integriert.

Betrieb: Regelmäßige Überprüfung der augmentierten Daten auf Konsistenz und Sinnhaftigkeit.

Risikobewertung: **Mittel**, da Datenaugmentation kann bei falscher Anwendung zu verzerrten Ergebnissen führen und es besteht das Risiko, dass das Modell aufgrund der augmentierten Daten zu spezifisch trainiert wird und dadurch in der Realität schlechtere Ergebnisse liefert.

10. Benchmark entwickeln: Es MUSS eine Benchmark-Bewertung für die Leistung neuer Modelle verwendet werden, um den Entwicklungsfortschritt zu messen und es SOLLTE hierbei ein Modell oder eine bewährte Metrik verwendet werden, um einen klaren Vergleich zu erhalten. Es DARF entsprechend angepasst werden, wenn sich

die Projektanforderungen ändern oder neue Erkenntnisse ermittelt werden, um keine Modellverzerrung zu erhalten.

Grundlegende Funktionalität: Ein Benchmark ist ein Vergleichspunkt, der als Basis für die Bewertung der Leistung eines Modells dient und es handelt sich dabei in der Regel um ein einfaches Modell oder eine festgelegte Leistungsmetrik, mit der zukünftige Modelle verglichen werden mit dem Ziel, den Fortschritt in der Modellentwicklung objektiv zu messen.

Vorgesehener Einsatzkontext: Benchmarks werden während des gesamten Entwicklungszyklus verwendet, um festzustellen, ob ein neues Modell eine Verbesserung gegenüber vorherigen Ansätzen darstellt und der Einsatz erfolgt sowohl in der Entwicklung als auch in der Evaluierung neuer Modelle.

Struktur der KI-Anwendung: Die Benchmark-Modelle können einfache Modelle oder bestehende Industriestandards sein, die als Referenz dienen, sowie die Leistung wird durch Metriken wie Genauigkeit, F1-Score, RMSE oder AUC bewertet sowie Benchmarks helfen dabei, die Effektivität neuer Ansätze zu objektivieren.

Maßnahmen:

Daten: Verwendung von standardisierten Datensätzen zur Evaluation, um Konsistenz und Vergleichbarkeit sicherzustellen.

KI-Komponente: Entwicklung eines Basis- oder Referenzmodells (z. B. ein einfaches lineares Modell), das als Maßstab dient.

Einbettung: Automatisierte Auswertung in der Modellpipeline, um die Leistung eines neuen Modells gegen die Benchmark zu testen.

Betrieb: Regelmäßige Überprüfung des Benchmarks, um sicherzustellen, dass er noch relevant ist und den Anforderungen des Projekts entspricht.

Risikobewertung: **Niedrig**, da Benchmarks nur als Vergleichspunkt dienen und keine unmittelbare Auswirkung auf die Produktion haben und es besteht jedoch das Risiko, dass die Benchmark zu einfach oder unpassend gewählt wird, was den Fortschritt irreführend darstellen könnte.

11. Mehrere KI-Modelle erstellen: Unterschiedliche Modelltypen, die entwickelt wurden, SOLLTEN verglichen werden, um das perfekte geeignete klassische Modell, neuronale Netze und somit den geeignetsten Algorithmus zu finden. Hierbei helfen automatisierte Tools (Grid Search, AutoML), um den Vergleich der Modelle zu erleichtern und die besten Hyperparameter zu finden.

Grundlegende Funktionalität: Die Entwicklung mehrerer KI-Modelle verfolgt das Ziel, durch Vergleich und Auswahl das am besten geeignete Modell für eine bestimmte Aufgabe zu finden sowie verschiedene Algorithmen (z. B. Entscheidungsbäume, Neuronale Netze, SVMs) werden getestet, um deren Vor- und Nachteile für das gegebene Problem zu verstehen und in manchen Fällen kann auch ein Ensemble von Modellen verwendet werden, um die Vorhersagegenauigkeit weiter zu steigern.

Vorgesehener Einsatzkontext: Dieser Schritt wird häufig zur Optimierung der Modellergebnisse eingesetzt, da der Kontext kann von der Suche nach dem besten Algorithmus für ein spezifisches Problem bis zur Verbesserung der Modellrobustheit durch den Einsatz von Ensemble-Methoden reichen.

Struktur der KI-Anwendung: Die Erstellung mehrerer Modelle erfordert ein Pipeline-Design, das es ermöglicht, verschiedene Algorithmen parallel oder sequentiell zu testen und automatisierte Frameworks wie AutoML oder Hyperparameter-Tuning-Plattformen helfen dabei, effizient das beste Modell auszuwählen.

Maßnahmen:

Daten: Sicherstellen, dass die Daten konsistent aufbereitet sind und für alle Modelle gleich verwendet werden sowie unterschiedliche Algorithmen können unterschiedlich empfindlich auf die Art der Vorverarbeitung reagieren.

KI-Komponente: Implementierung und Training verschiedener Modelltypen. Je nach Anwendung kann es sinnvoll sein, klassische Modelle wie Random Forests oder moderne neuronale Netze zu verwenden.

Einbettung: Der Modellvergleich wird durch Tools wie Grid Search oder Random Search durchgeführt. Frameworks wie scikit-learn oder Auto-ML helfen dabei, die besten Modelle effizient zu testen.

Betrieb: Das Modell, das die besten Ergebnisse liefert, wird für den produktiven Einsatz vorbereitet. Die anderen Modelle können als Backup oder alternative Ansätze weiter beobachtet werden.

Risikobewertung: Mittel, da das Risiko liegt in der Auswahl eines ungeeigneten Modells oder im Overfitting, wenn zu viele Modelle auf dieselben Daten trainiert werden sowie besteht die Gefahr, dass das beste Modell im Training gut abschneidet, aber in der Praxis schlechter performt.

12. Primäre Metriken evaluieren: Die Leistungsfähigkeit des Modells MUSS von spezifischen Metriken bewertet werden, um die wichtigen kritischen jeweiligen Anwendungen heraus zu filtern und somit SOLLTEN verschiedene Metriken im Vergleich stehen, um eine umfassende Modelleleistungsbewertung zu erhalten. Hierbei DÜRFEN Tools verwendet werden, wie F1-Score bei unausgewogenen Klassifikationsproblemen.

Grundlegende Funktionalität: Die Evaluierung primärer Metriken zielt darauf ab, die Leistung eines Modells basierend auf spezifischen Metriken zu bewerten, die für das Problem von zentraler Bedeutung sind sowie die Metriken wie Genauigkeit, Präzision, Recall, F1-Score, RMSE (bei Regressionen) oder AUC werden verwendet, um das Modell zu beurteilen mit dem Ziel ist es, die Modellgüte objektiv zu bewerten und zu entscheiden, ob das Modell bereit für den produktiven Einsatz ist.

Vorgesehener Einsatzkontext: Die Evaluierung erfolgt nach dem Training des Modells auf Validierungs- oder Testdaten sowie Metriken sind wichtig für die Entscheidung, ob das Modell in der Praxis erfolgreich eingesetzt werden kann oder ob weitere Verbesserungen notwendig sind.

Struktur der KI-Anwendung: Die Evaluierung der Metriken erfolgt nach dem Training des Modells und es ist wichtig, die Metriken so auszuwählen, dass sie das Ziel des Modells korrekt abbilden. Zum Beispiel ist bei einem Klassifikationsproblem mit unausgewogenen Daten der F1-Score oft relevanter als die reine Genauigkeit.

Maßnahmen:

Daten: Die Verwendung von Test- und Validierungsdaten, die nicht im Training verwendet wurden, ist essenziell, um eine unverzerrte Bewertung zu gewährleisten.

KI-Komponente: Das Modell wird anhand der definierten Metriken bewertet und dies hilft, die Stärken und Schwächen des Modells zu identifizieren.

Einbettung: Automatisierung der Metrik-Bewertung in der Trainingspipeline. Nach jedem Trainingslauf wird die Leistung des Modells gemessen und verglichen.

Betrieb: Überwachung der Metriken während des Betriebs, um sicherzustellen, dass die Leistung des Modells auch im Echtbetrieb stabil bleibt.

Risikobewertung: Niedrig, da die Bewertung der Metriken ist in der Regel risikofrei, solange geeignete Metriken gewählt wurden und kann jedoch zu Verzerrungen kommen, wenn die falschen Metriken verwendet werden, was die Leistung des Modells falsch darstellen kann.

13. Erklärbarkeit des KI-Modells: Die Modellsentscheidungen MUSS nachvollziehbar und erklärbar sein, um besonders sensible Aussagen transparenter für komplexer Modelle zu machen. Diese SOLLTEN mit Hilfe von Tools (LIME, SHAP) verwendet werden, um Prozesse zu vereinfachen, da ansonsten rechtliche und vertrauliche Probleme nicht erklärbar sind.

Grundlegende Funktionalität: Die Erklärbarkeit eines KI-Modells bezieht sich darauf, wie gut die Entscheidungen und Vorhersagen des Modells verständlich und nachvollziehbar gemacht werden können sowie in vielen Bereichen, wie z. B. der Medizin, dem Recht oder dem Finanzwesen, ist es wichtig, dass die Vorhersagen eines Modells nicht nur korrekt, sondern auch verständlich und erklärbar sind. Techniken wie LIME (Local Interpretable Model-Agnostic Explanations) oder SHAP (SHapley Additive exPlanations) werden eingesetzt, um komplexe Modelle transparenter zu machen.

Vorgesehener Einsatzkontext: Die Erklärbarkeit ist besonders in hochregulierten Bereichen notwendig, in denen Entscheidungen Auswirkungen auf Menschen oder Gesellschaften haben sowie ein weiteres Anwendungsgebiet ist die Erhöhung des Vertrauens der Benutzer in die KI-Entscheidungen, besonders bei Black-Box-Modellen wie neuronalen Netzen.

Struktur der KI-Anwendung: Die Erklärbarkeitstools werden entweder in das Modell selbst integriert (wie bei Entscheidungsbäumen) oder als externe Mechanismen hinzugefügt (wie bei Black-Box-Modellen und diese Tools helfen dabei, die Beiträge einzelner Merkmale zur Modellvorhersage zu verstehen.

Maßnahmen:

Daten: Bereitstellung von Daten, die verständlich gemacht werden können, um die Erklärbarkeit zu unterstützen und relevante Features für die Erklärbarkeit zu priorisieren.

KI-Komponente: Die Verwendung von erklärbaren Modellen oder spezifischen Werkzeugen, die Vorhersagen aufschlüsseln (z. B. SHAP oder LIME), ist zentral, um die Erklärbarkeit zu gewährleisten.

Einbettung: Die Erklärungen müssen für den Endnutzer verständlich präsentiert werden, z. B. durch Dashboards oder Visualisierungen, die die Entscheidungsfindung des Modells verdeutlichen.

Betrieb: Es ist wichtig, dass die Erklärungen nicht nur in der Entwicklungsphase, sondern auch im laufenden Betrieb leicht zugänglich sind sowie Schulungen für Endbenutzer können notwendig sein, um sicherzustellen, dass die Erklärungen richtig interpretiert werden.

Risikobewertung: Hoch, da Modelle, die nicht ausreichend erklärbar sind, stellen ein hohes Risiko darstellen, insbesondere in sensiblen oder hochregulierten Bereichen und unklare oder unverständliche Vorhersagen können das Vertrauen der Nutzer untergraben und zu rechtlichen oder ethischen Problemen führen.

Deploy AI/ML Engineer

14. Sekundäre Metriken evaluieren: Sekundäre Metriken MÜSSEN regelmäßig überprüft werden, um Hauptmetriken der Modelle und die Modelleleistung wie Genauigkeit, Fairness, Effizienz zu gewährleisten und dies kann mit Hilfe von Dashboards, erweiterte projektspezifische Metriken sein. Hierbei DÜRFEN NICHT die problematischen Bereiche wie Bias oder Energieverbrauch ignoriert werden.

Grundlegende Funktionalität: Sekundäre Metriken sind ergänzende Leistungskennzahlen, die neben den primären Metriken genutzt werden, um die Gesamtperformance eines KI-Modells besser zu verstehen sowie während Hauptmetriken wie „Accuracy“ oder „F1-Score“ die zentrale Leistungsfähigkeit eines Modells messen, liefern sekundäre Metriken zusätzliche Informationen über spezifische Aspekte, wie Robustheit, Fairness, Effizienz und Ausgewogenheit des Modells.

Einsatzkontext: In vielen Anwendungsbereichen, insbesondere in sensiblen Bereichen wie dem Gesundheitswesen, der Finanzindustrie oder bei Entscheidungen mit großen gesellschaftlichen Auswirkungen, reicht es nicht aus, sich nur auf die Hauptmetriken zu verlassen. Sekundäre Metriken spielen eine zentrale Rolle bei der Bewertung der Gesamtwirkung eines Modells und helfen, etwaige negative Nebeneffekte (z.B. Bias) zu identifizieren, die durch das Streben nach optimalen Hauptmetriken übersehen werden könnten.

Struktur der KI-Anwendung: Die Struktur von KI-Systemen, die sekundäre Metriken evaluieren, basiert auf erweiterten Bewertungsmechanismen sowie das Modell analysiert nicht nur seine Vorhersagegenauigkeit, sondern auch weiterführende Metriken wie „Fairness-Score“, „Resource-Usage“ (z.B. Rechenressourcen) oder „Stabilität über Zeit und integriert oft zusätzliche Module oder Mechanismen zur Überwachung der sekundären Metriken.

Maßnahmen:

Daten: Es ist wichtig, sicherzustellen, dass die Daten umfassend und repräsentativ sind sowie für die Berechnung sekundärer Metriken müssen die Daten z.B. detaillierte Informationen über sensible Attribute wie Geschlecht oder ethnische Zugehörigkeit enthalten, um Bias messen zu können und auch die Langzeitverfügbarkeit der Daten muss sichergestellt werden, um langfristige Trends und Effekte aufzuzeigen.

KI-Komponente: Das Modell muss in der Lage sein, mehrere Leistungsmetriken parallel zu evaluieren und neben der Hauptmetrik sollten auch sekundäre Metriken in die Bewertung einfließen, z.B. durch die Entwicklung von Multi-Objective-Optimierungsmodellen.

Einbettung: Es ist wichtig, dass die sekundären Metriken nahtlos in bestehende Systeme integriert werden und könnte durch Dashboards geschehen, die die verschiedenen Metriken visualisieren, oder durch regelmäßige Berichte, die aufzeigen, ob sich eine sekundäre Metrik in einem kritischen Bereich befindet.

Betrieb: Eine kontinuierliche Überwachung der sekundären Metriken sollte sichergestellt werden und in der Praxis bedeutet dies, dass bei jeder neuen Modellversion eine umfassende Prüfung aller relevanten Metriken durchgeführt wird, um zu verhindern, dass Verbesserungen in der Hauptmetrik negative Auswirkungen auf sekundäre Metriken haben.

Risikobewertung: Mittel, da bei der Evaluation sekundärer Metriken liegt darin, dass diese oft übersehen oder nicht ausreichend gewichtet werden und in Modell, das sich nur auf Hauptmetriken konzentriert, könnte in anderen wichtigen Bereichen wie Fairness oder Energieeffizienz versagen sowie sekundäre Metriken nicht korrekt ausgewertet oder priorisiert werden, kann dies in bestimmten Anwendungskontexten zu erheblichen Problemen führen, besonders bei Anwendungen mit hohen ethischen oder gesellschaftlichen Anforderungen.

15. KI-Modellentwicklung und Risikobewertung: Alle relevanten Risiken (z.B. Bias, Overfitting, ethische Bedenken, mögliche Verzerrungen, unvorhersehbares Verhalten) MUSS im Laufe der Modellentwicklung erkannt und kommuniziert sowie beseitigt werden. Hierbei SOLLTEN Risiken-Methoden reduziert werden (durch Cross-Validation, Bias-Überwachung), um die Modulzuverlässigkeit zu steigern.

Grundlegende Funktionalität: Die Entwicklung von KI-Modellen beinhaltet den gesamten Prozess von der Datenaufbereitung, dem Modelltraining bis hin zur Implementierung und während dieses Prozesses ist es essenziell, mögliche Risiken wie Verzerrungen (Bias), Überanpassung (Overfitting) oder unerwartetes Modellverhalten zu erkennen und zu bewerten und die Risikobewertung erfordert die Anwendung von Methoden zur Identifikation potenzieller Schwachstellen und ethischer Bedenken, die durch die Nutzung des Modells auftreten können.

Einsatzkontext: Besonders in Bereichen wie der Medizin, der Finanzindustrie oder in sicherheitskritischen Systemen muss die Entwicklung von KI-Modellen mit einer strengen Risikobewertung einhergehen und sind die Anforderungen an Genauigkeit, Fairness und Robustheit besonders hoch. Sowie fehlerhafte Modelle können nicht nur finanzielle Verluste, sondern auch ethische Konflikte oder gar körperliche Schäden verursachen.

Struktur der KI-Anwendung: Die Struktur umfasst oft komplexe, tief lernende Modelle (z.B. neuronale Netze), die durch kontinuierliches Training und Optimierung verbessert werden. Zusätzlich zur Modellentwicklung gehört die Implementierung von Mechanismen, die Risiken wie Modellversagen oder ethische Bedenken frühzeitig erkennen.

Maßnahmen:

Daten: Die Qualität und Repräsentativität der Daten ist entscheidend. Eine umfassende Analyse der Daten ist notwendig, um Verzerrungen (z.B. ungleiche Verteilung von Merkmalen wie Geschlecht oder ethnische Gruppen) frühzeitig zu identifizieren sowie Maßnahmen wie „Bias-Monitoring“ sollten während des gesamten Entwicklungsprozesses integriert werden.

KI-Komponente: Anwendung von Methoden wie Cross-Validation oder Bootstrapping zur Vermeidung von Überanpassung und Modelle sollten auf Robustheit gegenüber veränderten Eingabebedingungen getestet werden sowie müssen ethische Bedenken (z.B. in Bezug auf Datenschutz) adressiert werden.

Einbettung: Die Einbettung sollte ein Risikomanagement-Framework umfassen, das während der gesamten Modellentwicklung und des Einsatzes Risiken evaluiert und dokumentiert und regelmäßige Audits und Überprüfungen sind essenziell, um sicherzustellen, dass keine kritischen Risiken übersehen werden.

Betrieb: Während des Betriebs sollte eine kontinuierliche Risikobewertung stattfinden, die auch auf unvorhergesehene externe Faktoren reagieren kann, wie z.B. Veränderungen im

rechtlichen oder gesellschaftlichen Kontext und auch eine regelmäßige Reevaluation des Modells hinsichtlich seiner Performance und ethischen Implikationen.

Risikobewertung: Hoch, da in der Modellentwicklung und der Risikobewertung ist hoch, da viele unvorhersehbare Faktoren das Modellverhalten beeinflussen können sowie die Verzerrungen in den Daten oder das Versagen des Modells in unvorhergesehenen Szenarien stellen erhebliche Risiken darstellen und in sicherheitskritischen Bereichen kann dies zu massiven ethischen und finanziellen Problemen führen.

16. Nachentwicklungs-Überprüfung: Durch regelmäßige Modell-Leistungen und eventuelle Veränderungen MÜSSEN die Datenqualitätsveränderungen überprüft werden und es SOLLTEN alle Anomalien oder Performance-Abweichungen durch automatisierte Monitoring-Systeme sehr gut erkannt werden. Hierbei DÜRFEN die Module angepasst werden, um auf Datenveränderungen oder Einsatzkontext zu reagieren, Leistungsabfall und ungenaue Vorhersagen zu vermeiden.

Grundlegende Funktionalität: Die Nachentwicklungs-Überprüfung (Post-Development Review) ist ein zentraler Prozess, der sicherstellt, dass ein KI-Modell auch nach der ersten Implementierung weiterhin korrekt und effizient arbeitet und diese Überprüfung dient dazu, mögliche Performance-Verluste, Bias-Entwicklungen oder Veränderungen in der Datenqualität zu erkennen und zu beheben sowie ist notwendig, um die langfristige Einsatzfähigkeit und Verlässlichkeit eines Modells sicherzustellen.

Einsatzkontext: Die Nachentwicklungs-Überprüfung kommt in allen Bereichen zum Einsatz, in denen KI-Modelle nach der Entwicklung kontinuierlich genutzt werden und besonders bei dynamischen Umgebungen, in denen sich die Datenbasis oder externe Faktoren häufig ändern, ist diese Überprüfung essenziell, wie z.B. sind Anwendungsbereiche wie E-Commerce, Finanzmärkte oder Social Media, wo sich Nutzerverhalten oder Marktbedingungen rasch ändern können.

Struktur der KI-Anwendung: KI-Modelle in dieser Phase werden regelmäßig auf Anomalien und Performance-Veränderungen getestet und die Struktur kann MLOps-Prozesse integrieren, um ein kontinuierliches Monitoring und automatische Tests zu gewährleisten und diese Prozesse sind eng mit den operativen Systemen verknüpft und basieren oft auf Tools zur Drift-Erkennung, Modellvalidierung und Performance-Analyse.

Maßnahmen:

Daten: Kontinuierliche Datenüberwachung, um sicherzustellen, dass sich die zugrundeliegende Datenbasis nicht so verändert hat, dass das Modell ineffektiv wird (Daten-Drift) und die Daten sollten auf neue Anomalien und Trends überprüft und regelmäßig aktualisiert werden.

KI-Komponente: Automatische Validierungsmechanismen müssen implementiert werden, um das Modell in festgelegten Intervallen auf seine Performance zu überprüfen und Modellupdates sollten automatisiert durchgeführt werden, basierend auf diesen Analysen.

Einbettung: Ein robustes Überwachungssystem sollte nahtlos in die Produktionsumgebung integriert werden, das sowohl Performance-Metriken als auch sekundäre Metriken überwacht sowie solche Systeme könnten Warnungen auslösen, sobald eine signifikante Abweichung festgestellt wird.

Betrieb: Regelmäßige Modell-Updates und Revalidierung sind notwendig, um die langfristige Leistung sicherzustellen und proaktive Maßnahmen sollten ergriffen werden, um das Modell an neue Datenumgebungen oder Nutzungsanforderungen anzupassen.

Risikobewertung: Mittel, da bei der Nachentwicklungs-Überprüfung liegt in der Möglichkeit, dass Probleme erst zu spät erkannt werden, wenn Modelle nicht regelmäßig überprüft werden, können sie unter Umständen im Laufe der Zeit signifikante Performance-Verluste erleiden oder ethische Bedenken aufwerfen, insbesondere wenn sie sich in veränderten Datenumgebungen bewegen.

17. Operationalisierung mittels KI-Pipelines: Modellbereitstellung und -wartung über standardisierte Pipelines MUSS automatisiert werden, um kontinuierlich und reproduzierbare Ergebnisse zu garantieren. Hierbei SOLLTEN CI/CD-Strategien integriert werden, um die Pipeline-Automatisierung zu verbessern und den Modellaktualisierungsprozess vorantreiben. Dabei DÜRFEN die Pipeline-Struktur erneuert werden, um auf neue technische Anforderungen oder Datenquellen zu reagieren.

Grundlegende Funktionalität: Die Operationalisierung von KI-Modellen mittels Pipelines bedeutet die Automatisierung und Standardisierung des Prozesses, durch den KI-Modelle entwickelt, getestet und in Produktionsumgebungen bereitgestellt werden. KI-Pipelines ermöglichen es, diese Prozesse effizient, zuverlässig und reproduzierbar zu gestalten, indem sie Continuous Integration (CI) und Continuous Deployment (CD) nutzen.

Einsatzkontext: KI-Pipelines werden insbesondere in Unternehmen eingesetzt, die mehrere Modelle gleichzeitig verwalten oder schnell auf neue Anforderungen reagieren müssen und es ist besonders in dynamischen Umgebungen wie der Fertigung, der Telekommunikation und der Finanzindustrie wichtig, wo sich die Anforderungen an Modelle oder Daten schnell ändern können.

Struktur der KI-Anwendung: Eine KI-Pipeline umfasst mehrere Stufen – von der Datenvorverarbeitung, Modelltraining und -evaluierung bis hin zur Modellbereitstellung und ist typischerweise so strukturiert, dass sie einen klaren Fluss zwischen diesen Stufen gewährleistet, wobei jeder Schritt automatisiert und versioniert ist sowie CI/CD-Pipelines stellen sicher, dass bei Änderungen (z.B. neuen Daten oder Algorithmen) Modelle automatisch neu trainiert und getestet werden.

Maßnahmen:

Daten: Eine nahtlose Integration der Datenpipelines in die KI-Pipeline ist entscheidend und Daten müssen in Echtzeit verarbeitet und bereitgestellt werden, damit die Modelle kontinuierlich mit aktuellen Daten arbeiten können.

KI-Komponente: Die Modellbereitstellung muss automatisiert und versioniert erfolgen und es sollten Mechanismen zur Fehlererkennung eingebaut sein, die es ermöglichen, problematische Modellversionen schnell zu identifizieren und zurückzusetzen.

Einbettung: Die Pipeline sollte in bestehende IT-Systeme eingebettet sein, um eine nahtlose Interaktion zwischen Daten, Modellen und den operativen Systemen zu gewährleisten und dazu gehört auch die Integration von Monitoring- und Logging-Tools, die sicherstellen, dass jede Modellversion ordnungsgemäß funktioniert.

Betrieb: Eine regelmäßige Überwachung der gesamten Pipeline ist unerlässlich, um sicherzustellen, dass alle Prozessschritte wie geplant ablaufen sowie umfasst auch den Einsatz von Rückfallsystemen (Rollback), falls neue Modellversionen unerwartete Fehler aufweisen.

Risikobewertung: Mittel, da die Automatisierung mittels KI-Pipelines reduziert menschliche Fehler und beschleunigt die Prozesse, jedoch besteht das Risiko, dass Pipeline-Fehler auftreten, die schwer zu diagnostizieren sind und insbesondere bei komplexen Modellen oder Pipelines können unerwartete Fehler zu erheblichen Problemen führen, wenn nicht rechtzeitig geeignete Maßnahmen ergriffen werden.

18. Hyperautomatisierung von Prozessen und Systemen: Die Hyperautomatisierung MUSS stabil und robust integriert sein, um den störungsfreien Betrieb im automatisierten System zu gewährleisten. Hierbei SOLLTEN Überwachungsmechanismen und Fehlererkennung genutzt werden, um auf unvorhergesehene Probleme schnell reagieren zu können. Dabei DÜRFEN Korrekturen durchgeführt werden, um die Leistungssteigerungen und Automatisierungsanpassungen zu steigern mit manuellen Eingriffsmöglichkeiten, wenn die Einführung von neuen Datenquellen oder Technologien erfolgt.

Grundlegende Funktionalität: Hyperautomatisierung beschreibt den Einsatz von KI, Machine Learning und Robotic Process Automation (RPA), um Geschäftsprozesse vollständig zu automatisieren und menschliche Eingriffe zu minimieren und durch intelligente Systeme zu ersetzen, die Entscheidungen autonom treffen und Prozesse steuern sowie Hyperautomatisierung geht über die einfache Automatisierung hinaus, indem sie die Automatisierung in größerem Maßstab ermöglicht und verschiedene Technologien integriert, um Effizienz und Produktivität zu maximieren.

Einsatzkontext: Diese Technologie wird in Unternehmen verwendet, die einen hohen Automatisierungsgrad anstreben, z.B. in der Fertigung, im Kundenservice oder im Finanzwesen sowie repetitive Aufgaben, die auf klaren Regeln basieren, können durch Hyperautomatisierung ersetzt werden, um Skaleneffekte und Kosteneinsparungen zu erzielen.

Struktur der KI-Anwendung: Die Struktur der KI-Anwendung in der Hyperautomatisierung kombiniert verschiedene Technologien und umfasst RPA-Systeme, die mechanische Aufgaben automatisieren, und KI-Modelle, die komplexere Entscheidungen treffen sowie Systeme werden oft durch fortschrittliche Analysetools und Machine-Learning-Modelle ergänzt, die große Datenmengen verarbeiten und auf deren Basis Optimierungsentscheidungen treffen.

Maßnahmen:

Daten: Die Daten müssen vollständig strukturiert sein und in Echtzeit verarbeitet werden, um schnelle und genaue Entscheidungen zu ermöglichen sowie Datenspeicher- und Verwaltungssystem muss in der Lage sein, große Mengen an Informationen zu verarbeiten und zu analysieren, damit das Automatisierungssystem korrekt arbeiten kann.

KI-Komponente: KI-Modelle, die in der Hyperautomatisierung eingesetzt werden, müssen besonders robust und skalierbar sein und viele Entscheidungen ohne menschliche Überprüfung treffen, müssen sie in der Lage sein, Anomalien zu erkennen und selbstständig geeignete Maßnahmen zu ergreifen.

Einbettung: Die Einbettung von Hyperautomatisierungs-Systemen erfordert eine enge Integration in die IT-Infrastruktur eines Unternehmens, um sicherzustellen, dass alle Systeme miteinander kommunizieren und zusammenarbeiten können sowie umfasst auch die Sicherstellung, dass manuelle Eingriffe jederzeit möglich sind, falls das automatisierte System versagt.

Betrieb: Der Betrieb solcher Systeme erfordert eine fortlaufende Überwachung und Anpassung, um sicherzustellen, dass das System effizient arbeitet und müssen Wartungspläne und Rückfallsysteme implementiert werden, um das Risiko eines Systemversagens zu minimieren.

Risikobewertung: **Sehr hoch**, da bei der Hyperautomatisierung liegt in der vollständigen Abhängigkeit von automatisierten Systemen. Fehler in einem Hyperautomatisierungssystem können schwerwiegende Folgen haben, insbesondere wenn menschliche Eingriffe nicht schnell genug vorgenommen werden können sowie das System nicht flexibel genug ist, um auf unvorhergesehene Ereignisse zu reagieren, was zu ineffizienten Prozessen oder massiven Ausfällen führen kann.

Monitoring and Improvement vs. Deployment AI/ML Analyst

19. Leistung überwachen und bewerten: Eine kontinuierliche Überwachung der Modelleistung MUSS exakte und relevante Ergebnisse liefern, um weiterhin genaue und relevante Modellergebnisse zu liefern. Dabei SOLLTEN Dashboards und Alarmer verwendet werden, um Leistungsabweichungen in Echtzeit zu lokalisieren. Hierbei DÜRFEN Überwachungskriterien angepasst werden, wenn Anforderungen oder der Einsatzkontext geändert werden.

Grundlegende Funktionalität: Die Überwachung und Bewertung der Leistung von KI-Modellen ist ein kontinuierlicher Prozess, bei dem die Leistungsfähigkeit des Modells in der Produktion überwacht wird sowie sicherzustellen, dass das Modell dauerhaft den Anforderungen entspricht und keine signifikanten Leistungseinbußen erleidet und somit kommen Methoden zur Drift-Erkennung, Performance-Analyse und Echtzeitüberwachung zum Einsatz.

Einsatzkontext: In Unternehmen, die KI-Modelle für operative Zwecke einsetzen, ist die kontinuierliche Leistungsüberwachung unerlässlich sowie in Bereichen wie der Finanzanalyse, der Logistik oder der Online-Werbung kann ein Leistungsabfall eines Modells zu erheblichen finanziellen Verlusten oder ineffizienten Prozessen führen, wobei die Überwachung auch notwendig ist, um sicherzustellen, dass sich das Modell an sich ändernde Datenbedingungen anpasst.

Struktur der KI-Anwendung: Das KI-System ist mit einer Überwachungsinfrastruktur ausgestattet, die Echtzeitdaten sammelt und analysiert, um die Modelleistung zu bewerten sowie die Infrastruktur kann Anomalien erkennen und entsprechende Warnungen auslösen und

die Drift-Erkennungsmechanismen sorgen dafür, dass das Modell weiterhin korrekte Vorhersagen macht, auch wenn sich die zugrundeliegenden Daten ändern.

Maßnahmen:

Daten: Die Daten, die zur Überwachung verwendet werden, müssen aktuell und repräsentativ sein, um genaue Bewertungen der Modellleistung zu ermöglichen und muss ein System zur Erkennung von Daten-Drift implementiert werden, das Änderungen in der Datenverteilung frühzeitig erkennt.

KI-Komponente: Das Modell sollte mit Mechanismen zur Drift-Erkennung ausgestattet sein, um sicherzustellen, dass es bei Änderungen der Datenverteilung keine signifikanten Leistungseinbußen erleidet und sollten automatische Retests und Validierungen in regelmäßigen Abständen durchgeführt werden.

Einbettung: Ein Dashboard zur Echtzeitüberwachung sollte in das operative System eingebettet sein, um kontinuierlich Metriken wie Genauigkeit, Ausfallrate und Ressourcenverbrauch zu überwachen sowie sollten Protokolle und Alarme implementiert werden, die bei signifikanten Abweichungen ausgelöst werden.

Betrieb: Eine proaktive Wartung und Anpassung des Modells ist erforderlich, um auf Leistungseinbrüche zu reagieren und sollten regelmäßige Tests durchgeführt werden, um sicherzustellen, dass das Modell optimal funktioniert und bei Bedarf optimiert wird.

Risikobewertung: **Mittel**, da bei der Leistungsüberwachung besteht darin, dass ohne eine ausreichende Überwachung Leistungseinbußen oder Anomalien unbemerkt bleiben könnten und ein Modell, das nicht regelmäßig überwacht wird, könnte unter veränderten Bedingungen fehlerhafte Vorhersagen treffen, was zu signifikanten Geschäftsverlusten oder Fehlentscheidungen führen könnte.

20. Performance Monitoring: Wichtige Leistungsmetriken MÜSSEN (Genauigkeit, Latenzzeit und Ressourcenverbrauch) regelmäßig überwacht werden, um die Effizienz des Modells sicherzustellen. Es SOLLTEN automatisierte Berichtelogs und Warnungen verwendet werden, um Anomalien schnell zu erkennen und darauf Gegenmaßnahmen durchgeführt. Dabei DÜRFEN zusätzliche Überwachungsmetriken integriert werden, so dass Latenz- oder Ressourcenprobleme ausbleiben und die Anwendungskalierbarkeit zu garantieren.

Grundlegende Funktionalität: Performance Monitoring bezieht sich auf die kontinuierliche Überwachung der Modellleistung, um sicherzustellen, dass die KI auch nach der Bereitstellung in der Produktionsumgebung ihre Funktion optimal erfüllt und umfasst die Analyse von Metriken wie Genauigkeit, Präzision, Rückrufquote (Recall), F1-Score, Latenzzeiten, Durchsatz und andere relevante Leistungsindikatoren sowie sicherzustellen, dass das Modell die erwartete Leistung erbringt und keine Degradationen über die Zeit auftreten, insbesondere wenn sich die Eingabedaten oder Anwendungsbedingungen ändern. Durch Performance Monitoring können nicht nur technische Probleme wie langsame Reaktionszeiten oder ineffiziente Ressourcennutzung erkannt werden, sondern auch subtile Leistungsabfälle im Bereich der Vorhersagequalität, die durch Veränderungen in den zugrunde liegenden Daten oder Algorithmen verursacht werden sowie regelmäßige Erfassung von Metriken und deren Vergleich mit früheren Werten helfen dabei, die langfristige Konsistenz des Modells zu gewährleisten.

Vorgesehener Einsatzkontext: Performance Monitoring wird in praktisch allen KI-Anwendungen eingesetzt, besonders in solchen, die in produktiven Umgebungen laufen und kontinuierlich genutzt werden, z. B. Empfehlungssysteme, Betrugserkennungssysteme, autonome Fahrzeuge oder medizinische Diagnosesysteme sowie Kontexten ist es von entscheidender Bedeutung, dass das Modell auch unter realen Bedingungen zuverlässig arbeitet, um negative Auswirkungen auf Nutzer, Geschäftsprozesse oder das Gesamtsystem zu vermeiden.

Struktur der KI-Anwendung: Die Anwendung muss in der Lage sein, in Echtzeit oder in regelmäßigen Abständen Leistungsdaten zu erfassen und kann über integrierte Überwachungsdienste erfolgen, die eng mit der Produktionsumgebung verbunden sind sowie typischerweise ein Monitoring-System verwendet, das Alarme auslöst, wenn bestimmte Schwellenwerte überschritten werden, und Berichte generiert, die den Modellbetreuern eine detaillierte Einsicht in die Leistung und eventuelle Abweichungen geben. Performance Monitoring

erfordert also eine starke Kopplung zwischen dem Modell und der umgebenden Infrastruktur, um zeitnahe und relevante Daten zur Verfügung zu stellen.

Maßnahmen:

Daten: Implementieren Sie ein Datenerfassungs- und Speichersystem, das sowohl Eingabedaten als auch Modellvorhersagen protokolliert, um die Genauigkeit und die Vorhersagequalität zu überwachen.

KI-Komponente: Setzen Sie auf spezifische Leistungsmetriken, wie z. B. Präzision, Recall oder Latenzzeit, und erfassen Sie diese kontinuierlich während der Laufzeit des Modells.

Einbettung: Nutzen Sie ein Dashboard, das die kontinuierliche Sichtbarkeit der Modellleistung ermöglicht und sollte Alarme und visuelle Indikatoren enthalten, die Abweichungen von den erwarteten Leistungswerten anzeigen.

Betrieb: Entwickeln Sie einen Plan für regelmäßige Überprüfungen der Leistungsmetriken und führen Sie automatische Tests durch, um frühzeitig auf Performance-Veränderungen zu reagieren.

Risikobewertung: Mittel, da Fehlen eines robusten Performance-Monitorings kann dazu führen, dass Leistungseinbußen oder Datenanomalien zu spät bemerkt werden und könnte die Entscheidungsfindung des Modells beeinträchtigen und im schlimmsten Fall zu falschen Vorhersagen führen, die erhebliche Auswirkungen auf das Unternehmen oder die Nutzer haben.

21. Detect Data Drift: Die Drift-Erkennung in den Methoden MUSS integriert sein, um Datenverteilungsveränderungen frühzeitig zu erkennen und gegenzusteuern. Hierbei SOLLTEN regelmäßige Tests und andere Analyseverfahren durchgeführt werden, um die Konsistenz der Daten mit geeigneten Methoden sicherzustellen. Flache Vorhersagen und abweichende Genauigkeit der Modelle SOLLTEN vermeiden werden.

Grundlegende Funktionalität: Data Drift bezieht sich auf die Veränderung der Datenverteilung über die Zeit und die Verschiebungen können in den Eigenschaften der Eingabedaten auftreten (Covariate Drift), bei den Zielwerten (Label Drift) oder im zugrunde liegenden Zusammenhang zwischen Eingabe- und Ausgabedaten (Concept Drift). Es ist entscheidend, dass Methoden implementiert werden, um solche Drifts frühzeitig zu erkennen, da sie die Vorhersagegenauigkeit und die allgemeine Leistung des Modells signifikant beeinträchtigen können sowie das Erkennen von Data Drift ermöglicht es, Maßnahmen zu ergreifen, wie z. B. das Retraining des Modells oder die Anpassung der Datenaufbereitung, um die Leistung stabil zu halten.

Vorgesehener Einsatzkontext: Data Drift Monitoring ist besonders wichtig in Umgebungen, in denen die Datenströme oder -quellen dynamisch sind und sich die Eingabedaten über die Zeit ändern können, z.B. dafür sind Finanzmärkte, soziale Netzwerke oder Sensoren in IoT-Anwendungen, die ändern sich die Daten kontinuierlich, und Modelle, die auf veralteten oder sich verändernden Daten basieren, können schnell an Leistung verlieren, wenn keine Anpassungen vorgenommen werden.

Struktur der KI-Anwendung: Das System muss in der Lage sein, Veränderungen in der Datenverteilung automatisch zu erkennen und zu bewerten und werden häufig statistische Tests oder speziell trainierte Modelle zur Drift-Detektion verwendet sowie Komponenten sind eng in das Modell und die Datenpipeline integriert und führen regelmäßige Analysen durch, um Abweichungen von den ursprünglichen Trainingsdaten zu erkennen.

Maßnahmen:

Daten: Verwenden Sie statistische Methoden, wie z. B. den Kolmogorov-Smirnov-Test oder das Jensen-Shannon-Divergenzmaß, um Veränderungen in der Datenverteilung zu erkennen und implementieren Sie Machine-Learning-gestützte Techniken zur kontinuierlichen Überwachung von Datenströmen.

KI-Komponente: Entwickeln Sie Algorithmen, die selbstständig Data Drift erkennen und diese quantifizieren können und sollte automatisiert und regelmäßig durchgeführt werden sollten, um sicherzustellen, dass Modellanpassungen zeitnah erfolgen können.

Einbettung: Integrieren Sie Data-Drift-Detektionsmodule in die Produktionsinfrastruktur, um automatische Warnungen und Berichte bei signifikanten Änderungen der Datenverteilung zu erstellen.

Betrieb: Erstellen Sie Prozesse, die bei Erkennung von Data Drift automatische Modellanpassungen, Retraining oder sogar das Umschalten auf Notfallmodelle ermöglichen.

Risikobewertung: Hoch, da das Risiko eines unbemerkten Data Drifts ist erheblich, da dies zu fehlerhaften Vorhersagen und Entscheidungen führen kann und ein Modell, das auf veränderten oder nicht mehr repräsentativen Daten basiert, kann falsche oder schädliche Ergebnisse liefern, was schwerwiegende Konsequenzen für das Unternehmen oder die Nutzer haben kann.

22. Monitor for Anomalies: Anomalien von Eingabedaten und Modellvorhersagen MÜSSEN kontinuierlich überwacht werden, um frühzeitig potenzielle Probleme oder Sicherheitsrisiken zu ermitteln. Dabei SOLLTEN Überwachungen mit automatischen KI-Alarmlern und Berichte-Logs eingesetzt werden, um auf neue Bedrohungen oder Datenveränderungen in Echtzeit zu reagieren und Modellfehlern und Ausfallzeiten zu vermeiden.

Grundlegende Funktionalität: Die Überwachung auf Anomalien zielt darauf ab, ungewöhnliche Muster in den Eingabedaten oder Modellvorhersagen zu erkennen, die auf Fehler im Datenstrom, unerwartete Eingaben oder potenzielle Angriffe hindeuten könnten und besonders wichtig, um sicherzustellen, dass das Modell keine fehlerhaften oder unvorhersehbaren Entscheidungen auf der Grundlage von nicht repräsentativen oder manipulativen Daten trifft.

Vorgesehener Einsatzkontext: Anomalieüberwachung wird häufig in sicherheitskritischen Systemen oder Anwendungen eingesetzt, in denen fehlerhafte Entscheidungen zu schwerwiegenden Folgen führen können und dazu gehören autonome Systeme, medizinische Anwendungen, betrugspräventive Modelle oder kritische Infrastrukturen wie Energieversorgungsnetze sowie es unerlässlich, dass ungewöhnliche Eingaben oder Ergebnisse sofort erkannt und analysiert werden.

Struktur der KI-Anwendung: Die Anomalieerkennung erfordert ein zusätzliches Subsystem in der KI-Anwendung, das Eingabedaten und Modellvorhersagen überwacht sowie das Subsystem kann auf regelbasierten Systemen (z. B. Schwellenwerten) oder auf Machine-Learning-Methoden basieren, die darauf trainiert sind, unbekannte oder ungewöhnliche Datenmuster zu erkennen, aber auch die Anomalieüberwachung mit anderen Sicherheits- und Überwachungssystemen integriert, um verdächtige Ereignisse schnell zu melden.

Maßnahmen:

Daten: Entwickeln Sie Verfahren, um automatisch anomale Eingabedaten zu erkennen, z. B. durch statistische Ausreißeranalyse oder Machine-Learning-gestützte Anomalieerkennung.

KI-Komponente: Implementieren Sie Algorithmen zur Überwachung der Modellvorhersagen, die auf Unstimmigkeiten oder anomale Muster hinweisen können.

Einbettung: Integrieren Sie das Anomalieerkennungssystem in die Produktionsumgebung, sodass Alarmlern bei Erkennung von Anomalien sofort ausgelöst und analysiert werden können.

Betrieb: Etablieren Sie einen klaren Protokollablauf, der bei Erkennung von Anomalien automatische Maßnahmen wie das Abschalten des Modells, die Rückkehr zu einem sicheren Zustand oder das Eingreifen eines menschlichen Operateurs ermöglicht.

Risikobewertung: Mittel, da Anomalien können auf fehlerhafte Datenströme, interne Fehler oder potenzielle Angriffe hinweisen und einzelne Anomalien möglicherweise keine katastrophalen Auswirkungen haben, können sie im aggregierten Zustand signifikante Störungen verursachen, insbesondere in sicherheitskritischen Systemen.

23. Implement Feedback Loops: Feedback Loops MÜSSEN generell von Usern und neuen Datenquellen systematisch erfasst werden und diese fließen dann in den Entwicklungsprozess ein, um die Modellleistung zu steigern. Dabei SOLLTEN automatisierte Anpassungsprozesse eingegliedert werden. Hier DÜRFEN nach den Datenschutzgesetzen nicht automatisch erkannte Userfeedbacks auch manuelle ausgewertet werden, um spezifische Verbesserungen zu analysieren.

Grundlegende Funktionalität: Feedback-Loops ermöglichen es, Nutzerfeedback und neue Datenquellen kontinuierlich in das Modell einfließen zu lassen, um die Modellleistung zu ver-

bessern und es an neue Bedingungen anzupassen und ein solches System stellt sicher, dass das Modell mit den aktuellsten Daten arbeitet und sich an veränderte Benutzeranforderungen oder externe Faktoren anpasst.

Vorgesehener Einsatzkontext: Feedback-Loops sind besonders in Systemen sinnvoll, die dynamischen Veränderungen ausgesetzt sind, wie z. B. Empfehlungssysteme, Chatbots, oder interaktive Anwendungen sowie können Nutzerinteraktionen oder Echtzeitdaten verwendet werden, um das Modell ständig zu verbessern und es auf die Bedürfnisse der Nutzer auszurichten.

Struktur der KI-Anwendung: Die Anwendung muss Mechanismen enthalten, um Feedback zu erfassen, zu verarbeiten und in das System zurückzuführen und kann in Form von Nutzerinteraktionen, explizitem Feedback oder durch die Erfassung von Nutzungsdaten geschehen sowie das System sollte in der Lage sein, solche Daten in Echtzeit oder in regelmäßigen Intervallen zu verwenden, um das Modell zu aktualisieren oder neu zu trainieren.

Maßnahmen:

Daten: Sammeln Sie systematisch Nutzerfeedback und neue Eingabedaten, die zur Verbesserung des Modells genutzt werden können.

KI-Komponente: Entwickeln Sie Mechanismen zur effizienten Nutzung von Feedbackdaten, z. B. durch Verstärkungslernen oder inkrementelles Lernen.

Einbettung: Integrieren Sie Feedback-Schleifen in die Anwendung, sodass das Modell in Echtzeit oder periodisch an neue Daten und Benutzeranforderungen angepasst wird.

Betrieb: Etablieren Sie Prozesse, um gesammeltes Feedback regelmäßig zu bewerten und Anpassungen vorzunehmen, sei es durch Modellupdates oder durch direkte Eingriffe in die Architektur.

Risikobewertung: **Niedrig**, da das Fehlen von Feedback-Loops kann die kontinuierliche Verbesserung des Modells beeinträchtigen, stellt jedoch keine unmittelbare Gefahr dar und das Modell bleibt möglicherweise länger statisch, kann jedoch weiterhin funktionieren.

24. Secure the Model: Dabei MÜSSEN Sicherheitsmaßnahmen integriert werden, um das Modell vor Cyber-Angriffen und unbefugtem Zugriff entgegenzuwirken. Hierbei SOLLTEN regelmäßige Sicherheitsüberprüfungen und Penetrationstests sowie kontinuierliche Strategien nach BSI durchgeführt werden, um aktuelle Schwachstellen in Hard- und Software vor Datenschutzverletzungen und Manipulationen zu vermeiden.

Grundlegende Funktionalität: Die Sicherheit des KI-Modells ist von entscheidender Bedeutung, um das Modell vor äußeren Bedrohungen, wie Manipulation oder unbefugtem Zugriff, zu schützen. Sowie die Angriffe auf das Modell, wie z. B. adversarial attacks, können die Integrität der Modellvorhersagen gefährden und potenziell gefährliche oder betrügerische Ergebnisse erzeugen, aber auch viele KI-Modelle sind anfällig für Datenschutzverletzungen, insbesondere wenn sie mit sensiblen Daten arbeiten.

Vorgesehener Einsatzkontext: Sicherheitsmaßnahmen sind in allen sicherheitskritischen Anwendungen unerlässlich, z. B. in der Finanzwelt, der Cybersicherheit, im Gesundheitswesen oder in Systemen, die persönliche Daten verarbeiten sowie kann ein erfolgreicher Angriff erhebliche rechtliche und ethische Konsequenzen haben, da sensible Informationen gefährdet werden oder das Vertrauen in die Entscheidungen des Systems untergraben wird.

Struktur der KI-Anwendung: Die Sicherheitsarchitektur eines KI-Systems sollte umfassende Schutzmaßnahmen beinhalten, wie z. B. Authentifizierung, Verschlüsselung der Daten, Modellhärtung und die kontinuierliche Überwachung auf potenzielle Bedrohungen und besonders wichtig ist es, Techniken gegen adversarial attacks zu implementieren und sicherzustellen, dass das Modell auf verschiedene Angriffsszenarien vorbereitet ist sowie regelmäßige Sicherheitsupdates und Schwachstellenanalysen müssen Teil des laufenden Betriebs sein.

Maßnahmen:

Daten: Implementieren Sie strenge Verschlüsselungsprotokolle, um sicherzustellen, dass Daten während des Transports und der Speicherung sicher sind.

KI-Komponente: Verwenden Sie Methoden zur Härtung des Modells, wie adversarial training, um es gegen gezielte Angriffe zu schützen.

Einbettung: Integrieren Sie Sicherheits-Frameworks und Mechanismen, die das Modell vor externen und internen Bedrohungen schützen.

Betrieb: Führen Sie regelmäßige Audits und Penetrationstests durch, um potenzielle Schwachstellen zu identifizieren und zu beheben.

Risikobewertung: **Sehr hoch**, da ein Angriff auf das KI-Modell kann schwerwiegende Auswirkungen haben, von der Manipulation der Vorhersagen bis hin zu Datenschutzverletzungen sowie ohne kontinuierliche Sicherheitsüberwachung ist das Modell anfällig für bekannte und neu auftretende Bedrohungen.

25. Ensure Compliance: Die Einhaltung aller regulatorische Anforderungen und gesetzlichen Vorschriften von Industriestandards MÜSSEN vor sensiblen Daten sicherstellt werden. Es SOLLTEN regelmäßige Audits und Überprüfungen der Compliance geprüft werden, sodass das Modell den aktuellen Standards entspricht. Zusätzliche Maßnahmen DÜRFEN zur Erklärbarkeit und Transparenz des Modells integriert und angewandt werden, um gesetzliche Vorschriften gerecht zu werden und somit rechtliche und finanzielle Risiken zu vermeiden.

Grundlegende Funktionalität: Die Sicherstellung der Compliance bedeutet, dass das KI-Modell und seine Verwendung den geltenden rechtlichen Vorschriften und Industriestandards entsprechen und betrifft den Umgang mit Daten (insbesondere personenbezogenen Daten), die Erklärbarkeit und Nachvollziehbarkeit des Modells, sowie ethische Richtlinien und Regularien sowie KI-Modelle müssen transparent und nachvollziehbar sein, insbesondere wenn sie in regulierten Branchen wie dem Gesundheitswesen oder dem Finanzwesen eingesetzt werden.

Vorgesehener Einsatzkontext: Compliance-Maßnahmen sind in Branchen unerlässlich, die strengen gesetzlichen Vorgaben unterliegen, wie z. B. im Gesundheitswesen (HIPAA), der Finanzindustrie (GDPR, BaFin) oder dem öffentlichen Sektor und es geht darum, sicherzustellen, dass alle Aspekte des Modells von der Datenerfassung bis hin zur Entscheidungsfindung regelkonform sind.

Struktur der KI-Anwendung: Die Architektur der KI-Anwendung muss Mechanismen zur Einhaltung der relevanten Vorschriften und Standards enthalten und umfasst Verfahren zur Erklärbarkeit des Modells (z. B. XAI-Techniken), Auditing-Systeme, die die Entscheidungen des Modells nachvollziehbar machen, sowie Datenschutzmechanismen, die sicherstellen, dass personenbezogene Daten ordnungsgemäß verarbeitet und gespeichert werden.

Maßnahmen:

Daten: Implementieren Sie Datenschutzrichtlinien und setzen Sie Maßnahmen zur Einhaltung von Datenschutzbestimmungen (z. B. GDPR, HIPAA) um.

KI-Komponente: Entwickeln Sie erklärbare Modelle oder fügen Sie nachträgliche Erklärungsmechanismen hinzu, um die Entscheidungen des Modells nachvollziehbar zu machen.

Einbettung: Integrieren Sie Auditing-Mechanismen und regelmäßige Überprüfungen, um die Einhaltung von Vorschriften zu überprüfen.

Betrieb: Führen Sie regelmäßig Audits und Compliance-Prüfungen durch, um sicherzustellen, dass alle Prozesse und Entscheidungen den gesetzlichen und ethischen Standards entsprechen.

Risikobewertung: **Hoch**, da die Nichteinhaltung gesetzlicher Vorschriften kann zu erheblichen rechtlichen und finanziellen Konsequenzen führen, einschließlich Strafen und Klagen und kann ein Mangel an Compliance das Vertrauen der Nutzer und Stakeholder in die KI-Anwendung erheblich beeinträchtigen.

26. Update und Retrain: Modelle MÜSSEN regelmäßig aktualisiert und neu trainiert werden, um aktuelle rechtliche, langfristige und situationsabhängige Datenbedingungen zu garantieren und somit anerkannte Ergebnisse zu liefern. Hierbei SOLLTEN Prozesse für das Retraining automatisiert und durchgeführt werden, um Aufwände zu minimieren und die Effizienz zu steigern. Dabei DÜRFEN Modellarchitekturänderungen erstellt werden, wenn neue Daten oder Erkenntnisse aus der Wissenschaft erzielt wurden und somit Vernachlässigungen von Retrain-Prozess zu vermeiden bei veralteten Modellen, Ungenauigkeiten und Qualitätseinbußen negative zu beeinflussen.

Grundlegende Funktionalität: Das regelmäßige Update und Retraining von KI-Modellen

ist notwendig, um sicherzustellen, dass sie weiterhin valide und präzise Ergebnisse liefern. Modelle, die in Produktion sind, basieren auf historischen Daten und Annahmen, die sich im Laufe der Zeit ändern können und wenn Modelle veralten, kann ihre Leistung abnehmen – ein Phänomen, das als **Model Drift** bekannt ist sowie das Retraining eines Modells bedeutet, dass das Modell mit neuen Daten erneut trainiert wird, um die aktuellen Gegebenheiten besser abzubilden und dies stellt sicher, dass die Modellvorhersagen weiterhin genau und relevant bleiben. Dabei Updates können auch Änderungen an der Modellarchitektur oder den zugrundeliegenden Algorithmen beinhalten.

Vorgesehener Einsatzkontext: Die Notwendigkeit von Updates und Retraining tritt in allen Branchen auf, in denen sich die Datenlandschaft rasch ändern kann. Beispiele: **E-Commerce** mit Nutzerverhalten und Kaufpräferenzen ändern sich ständig, was regelmäßige Aktualisierungen der Empfehlungssysteme erfordert; **Finanzmärkte** mit Finanzmodellen müssen regelmäßig auf neue Markttrends und geopolitische Entwicklungen reagieren; **Social Media** mit Algorithmen zur Sentiment-Analyse müssen sich an neue Sprachtrends und Nutzerverhalten anpassen.

Struktur der KI-Anwendung: Der Retrain-Prozess beinhaltet: **Modellüberwachung** mit fortlaufende Überwachung der Modelleleistung, um festzustellen, wann ein Retraining notwendig ist; **Datenbeschaffung** mit Sammlung neuer Daten, die Veränderungen in der zugrundeliegenden Umgebung widerspiegeln; **Retraining** mit das Modell wird mit den neuen Daten erneut trainiert, um sicherzustellen, dass es sich an veränderte Bedingungen anpasst; **Evaluierung** mit das retrainierte Modell wird getestet, um sicherzustellen, dass es die angestrebten Leistungsziele erfüllt, bevor es in die Produktion überführt wird.

Maßnahmen:

Daten: Implementierung einer Pipeline für die kontinuierliche Erfassung und Anreicherung von Daten, die für das Retraining des Modells verwendet werden und Daten müssen regelmäßig auf Qualität und Aktualität geprüft werden.

KI-Komponente: Einführung von Mechanismen, die es ermöglichen, Modelle automatisch zu erkennen, die ein Retraining benötigen sowie sollten Verfahren entwickelt werden, um Modelle nach dem Retraining auf ihre Robustheit zu überprüfen.

Einbettung: Automatisierung des Retrain-Prozesses und der Modellbereitstellung, sodass Updates nahtlos in die Produktion integriert werden können und die Nutzung von DevOps-Tools zur Versionierung von Modellen.

Betrieb: Implementierung von Maßnahmen zur Risikominderung, um sicherzustellen, dass das retrainierte Modell keine negativen Auswirkungen auf das Gesamtsystem hat und kann durch umfangreiche Test- und Validierungsprozesse sowie durch den Einsatz von A/B-Tests erreicht werden.

Risikobewertung: Mittel bis Hoch, da das retrainierte Modell schlechter abschneidet als das ursprüngliche Modell oder dass unerwartete Interaktionen mit anderen Systemen auftreten und eine sorgfältige Planung und Überwachung des Update-Prozesses erforderlich, um negative Folgen zu vermeiden.

Hinweis: Diese Regeln bieten einen Rahmen, der sicherstellt, dass KI-Systeme verantwortungsvoll, sicher, fair und transparent entwickelt und eingesetzt werden. Weitere Regeln oder Unterteilungen können ergänzend erstellt werden!

Anhang III Glossar - Begriffserklärungen

Die digitale Transformation hat sich zu einem entscheidenden Element in der Arbeitswelt entwickelt. Die umfassende Vernetzung erstreckt sich über das Internet, das Internet der Dinge (IoT), die digitale Plattformökonomie, den mobilen E-Commerce und die Cloud-Transformation.



Abbildung 146: Übersicht von der Cyber-Security

Der Begriff Informationstechnik (IT) bezieht sich auf die digitale Verarbeitung von Daten. In diesem Kontext wird der Begriff IT-Sicherheit (oder auch Cyber-Security) verwendet. Die Funktionsfähigkeit (Safety) der Systeme sollte den gesetzlichen Vorgaben des Bundesamts für Sicherheit in der Informationstechnik (BSI) entsprechen. Die Informationssicherheit (Security) spielt dabei eine zentrale Rolle, indem sie sicherstellt, dass die technische Verarbeitung von Informationen durch funktionssichere Systeme gewährleistet wird. Nur autorisierte Personen sollten Zugriff auf Daten und Informationen in IT-Infrastrukturen haben. Dieses Thema gewinnt sowohl im privaten als auch im unternehmerischen Bereich zunehmend an Bedeutung, bedingt durch die steigende Nutzung von Hardware und Software. Ein Angriff auf die IT-Sicherheit ermöglicht es Angreifern, Informationen oder Daten zu erlangen, die nicht für Drittpersonen bestimmt sind. Dies kann durch veraltete IT-Systeme mit unsicheren Programmierungen, Manipulation der Mitarbeiter, Social Engineering, unzureichend geschützte mobile Endgeräte mit ungeschütztem Zugang Angriffe durchgeführt werden. Unternehmen sehen darin eine ernsthafte Gefahr für die Wirtschaft und die Gesellschaft weltweit, insbesondere in Zeiten wie in der Coronapandemie oder Kriegen, in denen Cyber-Angriffe stark zugenommen haben. Ein gesteigertes Bewusstsein für die kriminellen Energien der Angreifer hat dazu geführt, dass Unternehmen mehr in neue Sicherheitstechnologien und Infrastrukturen investieren müssen. Es ist notwendig, das Sicherheitsbewusstsein zu verbessern und Mitarbeiter zu schulen, um Angriffen entgegenzuwirken. Obwohl Systeme und Infrastrukturen nicht vollständig vor Angriffen geschützt werden können, bietet das Bundesamt für Sicherheit in der Informationstechnik (BSI) den Leitfaden der IT-Grundschutz-Methodik¹⁹⁸ an. Dieser enthält Maßnahmenkataloge, Risikofaktoren, Bewertungen, Notfallszenarien, Cyber-Warnungen, Sicherheitsrichtlinien sowie Schulungen für IT-Spezialisten. Beispielsweise kooperiert die Syss GmbH¹⁹⁹ mit IT-Sicherheitsspezialisten, um Systeme und Infrastrukturen zu testen und die Ergebnisse zu veröffentlichen. Das BSI bietet Unternehmen IT-Sicherheitsdienstleistungen und Schulungen für ihre Mitarbeiter an.

IT-Sicherheitsunternehmen bieten Unterstützung durch Penetrationstests und IT-Beratungen an, um Schwachstellen in Unternehmen aufzudecken und diese schnellstmöglich zu beheben und zu veröffentlichen. Auch Unternehmen, die Hardware und Software herstellen

vgl.¹⁹⁸ (BSI)
¹⁹⁹ (Syss.GmbH)

und entwickeln ein wachsendes Interesse und eine Sensibilität, um Schwachstellen in der IT-Infrastruktur von vornherein zu vermeiden. Hierbei kann zwischen physischen Schwachstellen, wie bei Diebstahl von Geräten oder Einbruch in IT-Bereiche, und natürlichen Schwachstellen, wie durch Umwelt oder umgebungsbedingte Gründe, unterschieden werden. Mögliche Ursachen für Schwachstellen können technische Fehler der Hardware oder Software, sicherheitsrelevante Programmierfehler (CERTs – Computer Emergency Response Team), Medien (Daten-träger), Vernetzungen (Netzwerksysteme) oder Fehlverhalten der Mitarbeiter (Social Engineering) sein.

nationale und internationale Standards IT-Sicherheit

Das deutsche National Coordination Centre for Cybersecurity (NKCS) fungiert als gemeinsame Kooperationsplattform von verschiedenen Ressorts, darunter das Bundesministerium für Wirtschaft und Energie (BMWi), das Bundesministerium des Innern (BMI), das Bundesministerium der Verteidigung (BMVg), das Bundesministerium für Bildung und Forschung (BMBF) sowie einzelner nachgeordneter Bereiche wie das Bundesamt für Sicherheit in der Informationstechnik (BSI), das Forschungsinstitut für öffentliche Verwaltung CODE (FI CODE) und die DLR Projektträger (DLR-PT). Die Gesamtkoordination liegt beim Bundesministerium des Innern (BMI). Das BSI übernimmt hierbei die zentrale Rolle als Kopfstelle ("Single Point of Contact") für das NKCS, sowohl im nationalen Kontext als auch für das europäische Netzwerk der nationalen Koordinierungszentren und für die Cybersicherheits-Community.²⁰⁰

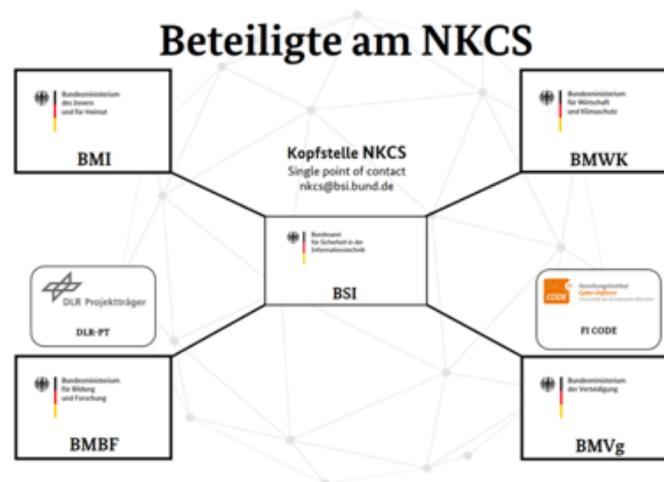


Abbildung 147: NKCS-Verbund

gesetzliche Grundlagen der IT-Sicherheit laut BSI

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) ist seit 1991 eine Bundesoberbehörde der Bundesrepublik Deutschland und untersteht dem Bundesministerium des Innern (BMI). Als zuständige Instanz für die IT-Sicherheit in der Informationsgesellschaft strebt das BSI danach, eine sichere Informations- und Kommunikationsgesellschaft in Deutschland zu gewährleisten und voranzutreiben. Die Zielgruppen umfassen öffentliche Verwaltungen, Wirtschaftsunternehmen, Wissenschafts- und Forschungseinrichtungen sowie Privatanwender. Auch ausländische Behörden, wie zum Beispiel die NSA, gehören dazu. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) übernimmt eine Vielzahl von Aufgaben, die im Zusammenhang mit der Gewährleistung der Informationssicherheit auf verschiedenen Ebenen stehen.

vgl.²⁰⁰ (BSI)

Im Folgenden werden die wichtigsten Funktionen des BSI detailliert ausformuliert.²⁰¹

Schutz der Netze des Bundes sowie Erkennung und Abwehr von Angriffen auf die Regierungsnetze: Das BSI ist verantwortlich für den umfassenden Schutz der Netze des Bundes. Diese Aufgabe beinhaltet nicht nur präventive Maßnahmen, sondern auch die kontinuierliche Überwachung, Erkennung und Abwehr von potenziellen Angriffen auf die sensiblen Regierungsnetze. Durch fortgeschrittene Sicherheitsmechanismen und reaktive Maßnahmen trägt das BSI dazu bei, die Integrität und Vertraulichkeit der digitalen Infrastruktur zu gewährleisten.

Prüfung, Zertifizierung und Akkreditierung von IT-Produkten und Dienstleistungen: Das BSI übernimmt die wichtige Aufgabe, IT-Produkte und -Dienstleistungen auf ihre Sicherheit hin zu überprüfen. Dies schließt die Entwicklung und Durchführung von Prüfverfahren ein, um sicherzustellen, dass die in Verwendung befindlichen Technologien den höchsten Sicherheitsstandards entsprechen. Durch die Vergabe von Zertifikaten und Akkreditierungen trägt das BSI dazu bei, dass nur vertrauenswürdige und sichere IT-Lösungen im Bundesbereich eingesetzt werden.

Warnung vor Schadprogrammen oder Sicherheitslücken in IT-Produkten und -Dienstleistungen: Das BSI fungiert als Frühwarnsystem, indem es aktiv Schadprogramme und Sicherheitslücken in verschiedenen IT-Produkten und -Dienstleistungen identifiziert. Durch die zeitnahe Warnung vor potenziellen Bedrohungen ermöglicht das BSI präventive Maßnahmen zur Minimierung von Risiken und Schäden.

IT-Sicherheitsberatung für die Bundesverwaltung und andere Zielgruppen: Das BSI bietet umfassende Beratungsdienste in Bezug auf IT-Sicherheit für die Bundesverwaltung sowie andere relevante Zielgruppen an. Diese Beratung erstreckt sich über verschiedenste Aspekte der Informationssicherheit und hilft den Organisationen, angemessene Sicherheitsstrategien zu entwickeln und umzusetzen.

Information und Sensibilisierung der Bürger für das Thema in der Informatik und Internet-IT-Sicherheit: Neben der Unterstützung von Institutionen legt das BSI großen Wert auf die Sensibilisierung der Bürger für Themen der IT- und Internet-Sicherheit. Hierzu werden Informationen bereitgestellt, die helfen, ein Bewusstsein für potenzielle Gefahren zu schaffen und die Bürger in die Lage versetzen, sicherere digitale Praktiken anzuwenden.

Entwicklung von einheitlichen und verbindlichen IT-Sicherheitsstandards: Das BSI hat eine führende Rolle bei der Entwicklung von einheitlichen sowie verbindlichen IT-Sicherheitsstandards. Diese Standards dienen als Leitlinien für die Implementierung von Sicherheitsmaßnahmen in verschiedenen Organisationen und tragen dazu bei, konsistente Sicherheitspraktiken zu etablieren.

Entwicklung von Kryptosystemen für die IT-Abteilung des Bundes: Im Rahmen seiner Aufgaben entwickelt das BSI fortschrittliche Kryptosysteme, die in der IT-Abteilung des Bundes zum Einsatz kommen. Diese Systeme gewährleisten eine sichere Verschlüsselung von sensiblen Daten und tragen somit maßgeblich zur Vertraulichkeit und Integrität digitaler Informationen bei.

Einsatz von Künstlicher Intelligenz: Als fortschreitende Technologie, auf Grund von wachsenden Datenmengen haben Künstliche Intelligenz (KI) und KI-Systeme in verschiedenen Anwendungsbereichen, einschließlich Textanalyse, Übersetzung und autonomes Fahren, an Bedeutung gewonnen. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) engagiert sich in der Grundlagenforschung, entwickelt Anforderungen und Prüfkriterien, um den sicheren Einsatz von KI zu gewährleisten.

Die Schnittstelle zwischen KI und IT-Sicherheit wird in drei Hauptbereichen erforscht: IT-Sicherheit für KI, IT-Sicherheit durch KI und Angriffe durch KI. Diese Themen stehen im Mittelpunkt von Forschungsbemühungen des BSI sowie internationaler Standardisierungsgremien. Das BSI betont die Bedeutung von Sicherheit, Robustheit und Nachvollziehbarkeit bei der Anwendung von KI. Es führt jährlich Workshops durch, um den aktuellen Stand der

vgl.²⁰¹ (BSI, 2024)ff.

Prüfbarkeit zu erfassen. Zudem untersucht es die Erklärbarkeit von KI und beschäftigt sich mit formalen Methoden zur Verifikation von KI-Modellen.

Entwicklung von Quantum-Machine-Learning (QML) bringt neue Sicherheitsaspekte mit sich: Dazu gehören Kryptographie mit QML und hier könnten herkömmliche kryptografische Verfahren gefährden, was eine Überarbeitung der Sicherheitsstandards erfordert; Datensicherheit mit Quantum-Machine-Learning könnte Datenschutz und Datensicherheit beeinträchtigen, weshalb robuste Schutzmechanismen erforderlich sind; Sicherheitsalgorithmen, um neue Sicherheitsalgorithmen, einschließlich post-quanten kryptografischer Verfahren, zu erforschen und zu entwickeln werden sowie Standards und Richtlinien vom BSI sollten für die sichere Entwicklung und Implementierung von QML-Technologien festgelegt werden. Die Sicherheitsaspekte müssen kontinuierlich überprüft werden, um mit den Fortschritten in Quantum-Machine-Learning Schritt zu halten.

Die Inhalte des IT-Grundschutzes des BSI umfassen die Beschreibung elementarer Gefährdungen in der Unternehmens-IT, allgemeine Anforderungen zum sicheren neuesten Stand der Technik sowie grundlegende Umsetzungshinweise und Empfehlungen. Diese ermöglichen einen reibungslosen und sicheren Betrieb des Unternehmens. Der IT-Grundschutz setzt sich dabei aus den BSI-Standards und dem IT-Grundschutz-Kompendium zusammen.²⁰²

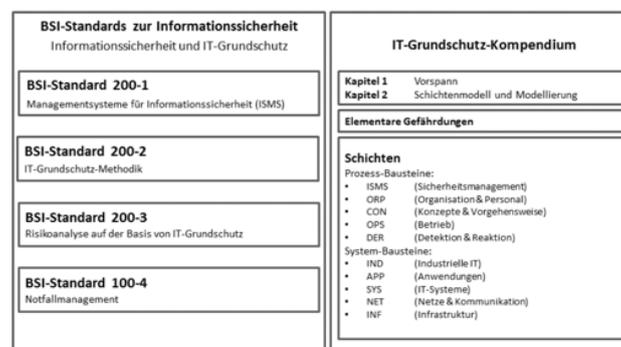


Abbildung 148: Übersicht über BSI-Publikationen zum Sicherheitsmanagement

Thema IS-Management nach ISO-27001

BSI-Standard ISO 27001 ist ein **Standard für Information Security Management System (ISMS)**, der in diesem Dokument, dem **ISO-200-1**, detaillierte Anforderungen festlegt. Der Standard legt fest, wie Informationssicherheit in einer Institution effektiv initiiert, gesteuert und überwacht werden kann. Er ist vollständig mit **ISO/IEC-27001** kompatibel und integriert die Begriffe sowie Empfehlungen aus **ISO/IEC-27000** und **ISO/IEC-27002**. Die vorliegende Norm bietet Lesern eine klare und systematische Anleitung, unabhängig von der gewählten Methode zur Umsetzung der Anforderungen an ein ISMS. Das BSI hat diesen Standard entwickelt, um den Inhalt der genannten ISO-Normen detaillierter zu behandeln und eine didaktisch verbesserte Darstellung zu ermöglichen. Die Gliederung wurde bewusst so gestaltet, dass sie mit der IT-Grundschutz-Vorgehensweise kompatibel ist. Damit wird eine präzise und umfassende Grundlage für die Implementierung von Informationssicherheit in Institutionen geschaffen.²⁰³

IT-Grundschutz-Methodik nach ISO-200-2

Die **IT-Grundschutz-Methodik** gemäß dem **BSI-Standard ISO 200-2** bietet einen umfassenden Leitfadens für den Aufbau und Betrieb eines Informationssicherheitsmanagementsystems in der Praxis. Zu den zentralen Themen gehören die Aufgaben des Informationssicherheitsmanagements sowie die Entwicklung einer geeigneten Organisationsstruktur für

vgl.²⁰² (BSI, 2023)

vgl.²⁰³ (BSI, 2023)

die Informationssicherheit. Die Methodik beschreibt detailliert, wie ein Sicherheitskonzept erstellt, passende Sicherheitsanforderungen ausgewählt und dessen Umsetzung systematisch durchgeführt werden kann. Sie beantwortet zudem die Frage, wie die Informationssicherheit im laufenden Betrieb aufrechterhalten und kontinuierlich verbessert werden kann. Um einen schrittweisen Einstieg in das Sicherheitsmanagement zu ermöglichen, bietet die Methodik unterschiedliche Ansätze, abhängig vom angestrebten Sicherheitsniveau und der Sensibilität der zu schützenden Informationen. Je nach bereits bestehenden Sicherheitsmaßnahmen in der Organisation kann es sinnvoll sein, vorübergehend von der vollständigen IT-Grundschutz-Vorgehensweise („Standard-Absicherung“) abzuweichen. Beispielsweise könnte eine Institution zunächst eine „Basis-Absicherung“ anstreben, indem sie die wichtigsten Sicherheitsanforderungen flächendeckend umsetzt, um schnell die größten Risiken zu minimieren. Alternativ könnte der Fokus auf den Schutz der besonders kritischen Werte der Organisation gelegt werden („Kern-Absicherung“).

Der IT-Grundschutz übersetzt die allgemeinen Anforderungen und Sicherheitsmaßnahmen der **ISO-Normen 27001** und der **ISO-Normen 27002** praxisnah auf Basis der **ISO-Normen 200-2**. Er unterstützt Anwender durch ausführliche Hinweise, Hintergrundinformationen und konkrete Beispiele bei der Umsetzung. Die Bausteine des IT-Grundschutz-Kompendiums erklären, was zu tun ist, während die Umsetzungshinweise klare Anleitungen bieten, wie diese Anforderungen – auch auf technischer Ebene – realisiert werden können. Das Vorgehen nach IT-Grundschutz stellt somit eine bewährte und effiziente Methode dar, um den Anforderungen der relevanten ISO-Normen umfassend gerecht zu werden.²⁰⁴

Risikoanalyse auf der Basis von IT-Grundschutz nach ISO-200-3

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) hat eine spezifische Methodik für die **Risikoanalyse** entwickelt, basierend auf dem IT-Grundschutz. Der **BSI-Standard ISO-200-3** legt dar, wie eine vereinfachte Risikoanalyse für die Informationsverarbeitung durchgeführt werden kann, aufbauend auf der etablierten IT-Grundschutz-Vorgehensweise. Diese Analyse basiert auf den grundlegenden Gefährdungen, die im IT-Grundschutz-Kompendium ausführlich beschrieben sind und als Grundlage für die Erstellung der IT-Grundschutz-Bausteine dienen. Dieses Vorgehen eignet sich besonders für Unternehmen oder Behörden, die bereits erfolgreich mit dem IT-Grundschutz arbeiten und eine Risikoanalyse nahtlos an ihre bestehende IT-Grundschutz-Analyse anknüpfen möchten. Indem sie auf den bewährten Grundlagen des IT-Grundschutzes aufbaut, ermöglicht die Methodik eine Effiziente und gut integrierte Risikobewertung für die Informationssicherheit. Dieser Ansatz bietet eine praktische Möglichkeit, die Risikoanalyse in bestehende Sicherheitspraktiken zu integrieren und somit einen umfassenden Schutz der Informationsverarbeitung zu gewährleisten.²⁰⁵

Notfallmanagement nach ISO-100-4

Der **BSI-Standard ISO-100-4** präsentiert eine methodische Herangehensweise zur Einführung und kontinuierlichen Aufrechterhaltung eines **Notfallmanagements** auf behördlicher oder unternehmensweiter Ebene. Die dargelegte Methodik baut auf der im **ISO-200-2** beschriebenen IT-Grundschutz-Vorgehensweise "Standard-Absicherung" auf und erweitert diese sinnvoll mit folgende Aufgaben:²⁰⁶ **Risikobewertung und -analyse**, welche Identifizierung und Bewertung potenzieller Notfallszenarien durchführt, um die wichtigsten Bedrohungen für die Organisation zu erkennen und entsprechende Maßnahmen zu planen; **Notfallplanung**, die Entwicklung detaillierter Notfallpläne, die konkrete Maßnahmen zur Bewältigung identifizierter Risiken enthalten sowie umfasst die Ressourcenplanung und die Koordination mit externen Parteien wie Rettungsdiensten und Behörden; **Implementierung und**

vgl.²⁰⁴ (BSI, 2023)

vgl.²⁰⁵ (BSI, 2023)

vgl.²⁰⁶ (BSI, 2023)

Betrieb, wie Schulung der Mitarbeiter für Notfälle und Durchführung regelmäßiger Übungen, um die Wirksamkeit der Pläne zu überprüfen und zudem werden Kommunikationssysteme für den Notfallbetrieb eingerichtet; **Überwachung und Bewertung** durch kontinuierliche Überwachung der Risikolandschaft und Bewertung der Notfallpläne, um ihre Wirksamkeit zu gewährleisten und notwendige Anpassungen vorzunehmen; **kontinuierliche Verbesserung** durch Analyse von Erfahrungen aus Notfällen und Übungen zur kontinuierlichen Verbesserung der Notfallpläne und Anpassung an neue Erkenntnisse und Umstände; **Managementbewertung**, um regelmäßige Überprüfungen durch das Management zur Sicherstellung der Angemessenheit, Effektivität und Effizienz des Notfallmanagementsystems, einschließlich strategischer Anpassungen bei Bedarf. Dieser strukturierte Ansatz hilft Organisationen, Notfälle systematisch zu bewältigen und die Sicherheit von Mitarbeitern und Vermögenswerten zu gewährleisten, wodurch die Widerstandsfähigkeit gegenüber unerwarteten Ereignissen gestärkt wird.

IT-Grundschutz - Grundpfeiler der Informationssicherheit

Die Sicherung von Informationen ist von grundlegender Bedeutung für Unternehmen und Behörden, da diese einen erheblichen Wert repräsentieren. In einer digitalisierten Welt, in der Geschäftsprozesse stark von IT abhängig sind, ist eine zuverlässige Informationsverarbeitung entscheidend für den reibungslosen Betrieb. Unzureichend geschützte Informationen stellen ein unterschätztes, potenziell existenzbedrohendes Risiko dar. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) bietet mit dem IT-Grundschutz eine praxisorientierte Methode, um Information angemessen zu schützen. Durch die Kombination von Basis-, Kern- und Standard-Absicherung sowie dem IT-Grundschutz-Kompendium bietet es Sicherheitsrichtlinien für den Aufbau eines ISMS.

Der IT-Grundschutz kann sowohl von kleinen und mittleren Unternehmen (KMU) als auch von großen Institutionen zur Implementierung eines ISMS genutzt werden. Eine erfolgreiche Umsetzung erfordert die Etablierung einer Organisationsstruktur (IT-Betrieb), die für die interne IT-Infrastruktur Einrichtung, Überwachung und Wartung verantwortlich ist. Infolge von Sicherheitsvorfällen besteht die Gefahr, Imageschäden zu erleiden. Daher ist es entscheidend, Daten angemessen zu schützen, Sicherheitsmaßnahmen sorgfältig zu planen, umzusetzen und zu kontrollieren. Informationssicherheit erfordert eine ganzheitliche Betrachtung und ist stark von infrastrukturellen, organisatorischen und personellen Rahmenbedingungen abhängig. Aspekte wie die Sicherheit der Betriebsumgebung, Schulung der Mitarbeiter, Zuverlässigkeit von Dienstleistungen und der korrekte Umgang mit Informationen dürfen nicht vernachlässigt werden. Schwächen in der Informationssicherheit können schwerwiegende Konsequenzen nach sich ziehen und potenziell Schäden in den Bereichen Verfügbarkeit, Vertraulichkeit und Integrität von Informationen verursachen. Der Verlust in einer dieser Kategorien kann die Funktionsfähigkeit und den Schutz sensibler Daten erheblich beeinträchtigen.

Aktuelle Entwicklungen in der Informations- und Kommunikationstechnik

Die **Information and Communication Technology (ICT)** spielt eine entscheidende Rolle im täglichen Leben, wobei Innovationen weiterhin rasch voranschreiten. Hervorzuhebende Entwicklungen sind.

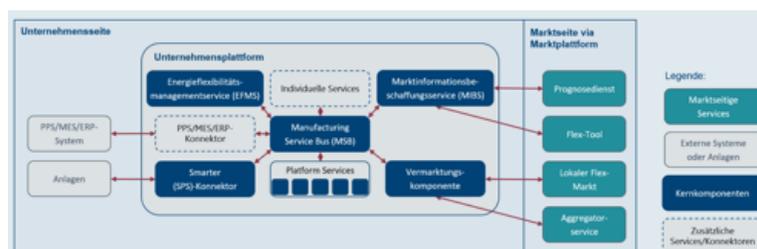


Abbildung 149: Aufbau der Energiesynchronisationsplattform zur Automatisierung und Standardisierung des Energieflexibilitätshandels

Steigender Vernetzungsgrad durch zunehmende Abhängigkeit von vernetzten IT-Systemen und Datennetzen mit potenziell globalen Auswirkungen.

IT-Verbreitung und Durchdringung durch Integration von IT in verschiedene Lebensbereiche mit immer kleineren und kostengünstigeren Hardwareeinheiten.

Verschwinden der Netzgrenzen durch Auflösung von Grenzen durch Cloud-Dienste und Internetkommunikation.

Kürzere Angriffszyklen durch Notwendigkeit eines effektiven Informationssicherheitsmanagements aufgrund der kurzen Zeitspanne zwischen Sicherheitslückenbekanntmachung und Angriffen.

Höhere Interaktivität von Anwendungen durch Kombination vorhandener Technologien für neue Anwendungsmodelle führt jedoch zu höherer Komplexität.

Verantwortung der Benutzer durch Betonung der Bedeutung des verantwortungsbewussten Handelns der Benutzer für effektive Informationssicherheit.

Die Berücksichtigung dieser Entwicklungen in der Informationssicherheitsstrategie ist unerlässlich, um den wachsenden Herausforderungen effektiv zu begegnen.²⁰⁷

Weitere Sicherheitsstandards COBIT 5

Control Objectives for Information and Related Technologies 5 (COBIT 5) betrachtet, laut BSI Grundschatz, die IT als eine entscheidende Grundlage für den Erfolg einer Institution bei der Verwirklichung ihrer Geschäftsziele. Der Rahmen fordert, dass die Ziele der IT aus der übergeordneten Geschäftsstrategie abgeleitet werden und die bereitgestellten Services den Qualitätsanforderungen der Geschäftsprozesse entsprechen. Ähnlich wie Information Technology Infrastructure Library (ITIL) setzt auch COBIT 5 auf zielgerichtete und optimierte IT-Prozesse. Ein innovativer Aspekt von COBIT 5 ist die Einführung des Prozesspotenzials, der die Fähigkeit einer Institution bewertet, die definierten Ziele zuverlässig und nachhaltig zu erreichen.

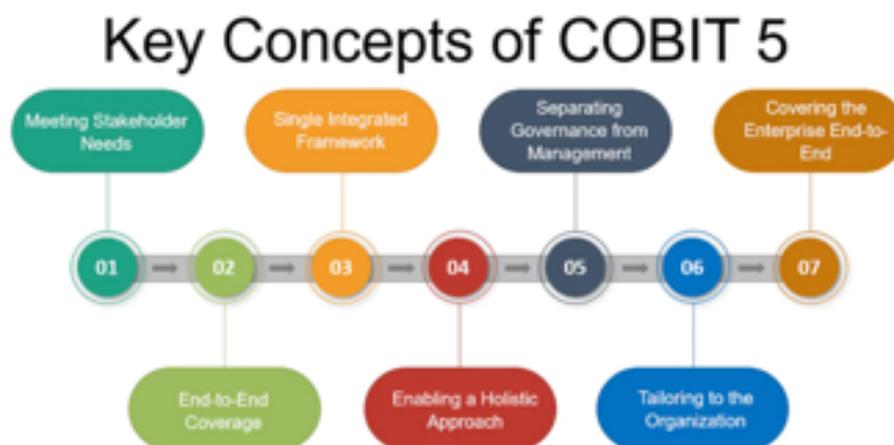


Abbildung 150: Key Concepts of COBIT 5

Die Professionalität der unterstützenden IT-Prozesse kann durch eine umfassende Bewertung der Reife aller 37 Prozessbereiche ermittelt werden, die in fünf Domänen unterteilt sind. Die COBIT-Richtlinien werden von der Information Systems Audit and Control Association (ISACA) herausgegeben. Bei der Entwicklung von COBIT orientierten sich die Autoren an etablierten Normen und Standards im Bereich des Sicherheitsmanagements, insbesondere an der **ISO/IEC-27002**.²⁰⁸

vgl.²⁰⁷ (BSI, 2023)
vgl.²⁰⁸ (BSI, 2023)

ITIL

Die **IT Infrastructure Library (ITIL)** stellt eine umfassende Literatursammlung im Bereich des "IT-Service-Management" dar und wurde unter der Leitung des britischen Office of Government Commerce (OGC) entwickelt. Laut der Definition von BSI fokussiert ITIL sich auf das Management von IT-Services aus der Perspektive des IT-Dienstleisters, der sowohl eine interne IT-Abteilung als auch ein externer Service Provider sein kann. Das übergeordnete Ziel besteht darin, die Qualität von IT-Dienstleistungen zu optimieren und die Kosteneffizienz zu verbessern. Die Informationssicherheit wird im Rahmen dieser Services aus einer operativen Perspektive analysiert. In umgekehrter Weise bildet ein reibungsloser IT-Betrieb einen entscheidenden Stützpfeiler für das Informationssicherheitsmanagementsystem (ISMS). Daher finden sich viele Prinzipien der ITIL auf ähnliche Weise wieder, wobei der Fokus deutlich auf Informationssicherheit im Kontext des IT-Grundschutzes und anderer Sicherheitsstandards liegt. Basierend auf den Prinzipien der ITIL wurde die Norm ISO/IEC-20000 entwickelt, die als Grundlage für die Zertifizierung von Service-Management-Systemen dient.²⁰⁹



Abbildung 151: ITIL-Zertifizierung von Service-Management-Systemen

NIST

Das **National Institute of Standards and Technology (NIST)** ist eine US-amerikanische Bundesbehörde, die unter anderem für die Entwicklung verbindlicher Standards zuständig ist, die für staatliche Einrichtungen in den USA verpflichtend sind. In der „Special Publication 800“-Serie (NIST SP 800) veröffentlicht die NIST regelmäßig Richtlinien zu verschiedenen Aspekten der Informationssicherheit, wie etwa Kryptografie und Cloud-Computing. Diese Dokumente sind nicht nur eine wertvolle Informationsquelle, sondern beeinflussen auch weltweit die Gestaltung von Informationssicherheitsstandards maßgeblich.



Abbildung 152: NIST Cyber Security Framework 2.0

vgl.²⁰⁹ (BSI, 2023)

Ein herausragendes Beispiel ist das **Dokument NIST SP 800-53** mit dem Titel "Security and Privacy Controls for Federal Information Systems and Organizations". Es stellt eine umfassende Sammlung von Sicherheitskontrollen für den Bereich Sicherheitsmanagement dar, die dazu dienen, Informationsverbindungen zu schützen. Die Kontrollen sind thematisch in verschiedene Bereiche unterteilt, darunter Schulung und Sensibilisierung sowie Berechtigungsmanagement und Infrastruktursicherheit.²¹⁰

ISF – The Standard of Good Practice

Das **Information Security Forum (ISF)** ist eine unabhängige, weltweit tätige Organisation, die sich auf Informationssicherheit konzentriert. Mit dem „Standard of Good Practice“ (SoGP) veröffentlicht das ISF einen praxisorientierten Leitfaden, der bewährte Verfahren und Best Practices zusammenführt. Dieser Leitfaden ist so gestaltet, dass er die Anforderungen zentraler Standards wie ISO/IEC-27002, COBIT 5, PCI DSS 3.1 und dem NIST Cybersecurity Framework abdeckt. Der SoGP unterteilt die behandelten Themen in verschiedene Bereiche, darunter Sicherheitsgovernance und Informationsrisikobewertung, und bietet Unternehmen wertvolle Unterstützung bei der Entwicklung und Umsetzung wirksamer Sicherheitsmaßnahmen.²¹¹

DISA

Die **Defense Information Systems Agency (DISA)** ist in verschiedenen militärischen Belangen tätig, darunter die Errichtung von Kommunikationsnetzwerken in Krisen- und Kriegsgebieten. Innerhalb der DISA gibt es einen spezialisierten Bereich, der sich auf "Cyber Security" konzentriert. In diesem Kontext veröffentlicht die Defense Information Systems Agency die DISA Security Technical Implementation Guides (STIGs). Diese Guides ermöglichen die Durchführung einer umfassenden Systemhärtung für Windows 10, wodurch die Sicherheit Ihrer Systeme gegenüber potenziellen Angriffen signifikant verbessert werden kann.²¹²

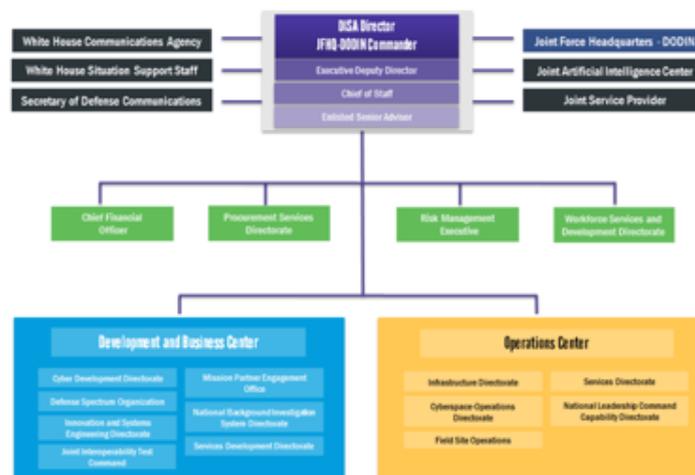


Abbildung 153: DISA-Überblick

CIS

Center for Internet Security (CIS) ist eine unabhängige Nonprofit-Organisation, der sowohl Einzelpersonen als auch Unternehmen beitreten können. Die Mission dieser Organisation lautet: *"Unser Ziel ist es, die vernetzte Welt sicherer zu gestalten, indem wir zeitgemäße Best-Practice-Lösungen entwickeln, validieren und fördern. Diese Lösungen sollen Menschen, Unternehmen und Regierungen dabei unterstützen, sich vor allgegenwärtigen*

vgl.²¹⁰ (BSI, 2023)

vgl.²¹¹ (BSI, 2023)

vgl.²¹² (FB Pro GmbH)

Cyber-Bedrohungen zu schützen.²¹³ Das Center for Internet Security veröffentlicht verschiedene Richtlinien und Tools, darunter die CIS Benchmarks, das CIS-CAT Pro und die CIS Hardened Images. Mithilfe dieser Vorgaben und Werkzeuge ist es möglich, Systeme wie den Windows Server zu stärken und deren Sicherheit gegenüber Cyber-Bedrohungen zu erhöhen.²¹⁴



Abbildung 154: CIS

ACSC



Abbildung 155: ACSC-Logo

In Australien operiert die IT-Sicherheitsbehörde unter dem Namen **Australian Cyber Security Centre (ACSC)**. Diese Institution wurde 2014 als Nachfolger des Cyber Security Operations Centre gegründet und ist der Australian Security Intelligence Organisation (ASIO) sowie dem Minister for Defence, dem Verteidigungsministerium, unterstellt. ASIO hat die Aufgabe, Cyber-Angriffe zu untersuchen und Gegenmaßnahmen zu entwickeln. Analog zu vergleichbaren Institutionen wie dem BSI in Deutschland gibt das ACSC umfassende Empfehlungen und Ratschläge für Unternehmen heraus. Ziel ist es, diese bei der Sicherung ihrer Systeme und IT-Infrastrukturen zu unterstützen. Dies schließt auch detaillierte Richtlinien für die Verstärkung von Systemen, sogenannte Hardening Guidelines, mit ein. Die Relevanz der Informationssicherheit nimmt in signifikantem Maße zu. Durch die voranschreitende Digitalisierung besteht die potenzielle Gefahr, dass Hacker und andere Cyberkriminelle im schlimmsten Fall auf sensible Daten von Unternehmen und Bürgern zugreifen können. Dies gilt es mit höchster Dringlichkeit zu verhindern.

Einige Organisationen wie das BSI, die DISA, das ACSC, das CIS und ähnliche Einrichtungen unterstützen Regierungen und Unternehmen dabei, Bedrohungen zu erkennen und mögliche Schäden zu minimieren. Eine von zahlreichen Maßnahmen in diesem Kontext besteht in der Intensivierung der Sicherheitsvorkehrungen von IT-Infrastrukturen sowie einzelner Systeme.²¹⁵

Schutzziele in der Informationssicherheit

Gemäß dem BSI gibt es drei grundlegende **Schutzziele** in der Informationssicherheit: **Vertraulichkeit**, **Verfügbarkeit** und **Integrität**. Diese werden durch Aspekte wie Authentisierung, Autorisierung, Datenschutz, Datensicherheit, Datensicherung, Penetrationstests, Risikoanalyse und Sicherheitsrichtlinien erweitert.

²¹³ (FB Pro GmbH)
²¹⁴ (FB Pro GmbH)
²¹⁵ (FB Pro GmbH)

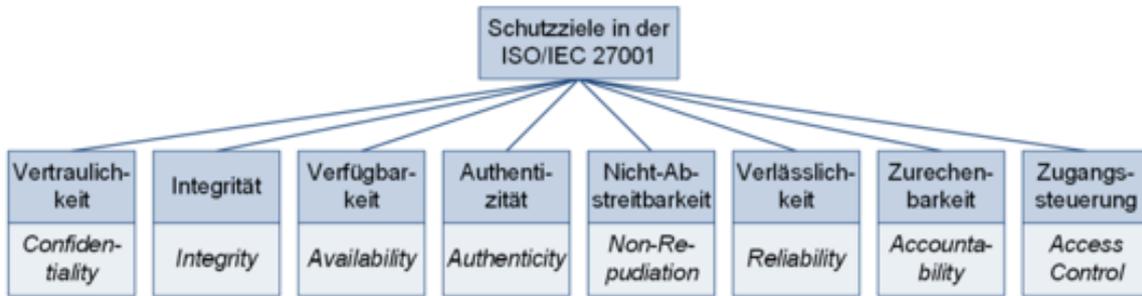


Abbildung 156: BSI-Schutzziele ISO/IEC 27001

Die **Schutzziele nach dem BSI** sind wie folgt:

Vertraulichkeit: Schutz vor unbefugter Nutzung vertraulicher Informationen und sensibler Daten.

Verfügbarkeit: Bereitstellung von Dienstleistungen, Funktionalitäten und Informationen durch IT-Systeme oder IT-Anwendungen.

Integrität: Sicherstellung, dass Daten und Informationen vollständig und unverändert dem IT-System zur Verfügung gestellt werden.

Authentisierung: Überprüfung und Bestätigung der Identität von Personen und IT-Systemen bei Anmeldungen.

Autorisierung: Überprüfung der Zugriffsrechte von Personen auf verschiedene Anwendungen.

Datenschutz: Schutz personenbezogener Daten vor unbefugtem Zugriff.

Datensicherung: Erstellung von Sicherungskopien, um Datenverlust zu verhindern.

Penetrationstests: Gezielte Tests zur Identifikation von Schwachstellen in IT-Systemen.

Risikoanalyse: Untersuchung möglicher Schäden und deren potenzielle Auswirkungen.

Sicherheitsrichtlinien: Festlegung von Schutzzielen und -maßnahmen durch Unternehmen oder Behörden.²¹⁶

Schadensszenarien

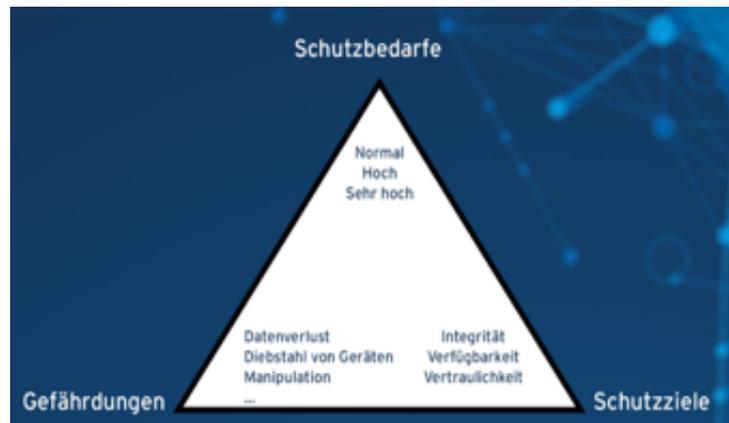


Abbildung 157: Gefährdungen - Schutzziele - Schutzbedarfe

Schritt 1: Festlegung der Schutzbedarfe liegt der Fokus auf der Präzisierung der Schutzbedarfe. Das Ziel besteht darin, die Daten, die Vertraulichkeit, Integrität und Verfügbarkeit erfordern, im Zeitpunkt eines möglichen Schadensfalls zu identifizieren. Dabei werden die Auswirkungen auf die betroffene Anwendung, Infrastruktur oder Architektur berücksichtigt. Zur Strukturierung dieser Schutzbedarfe verwendet das BSI Kategorien, die in einer festgelegten Reihenfolge Schadensszenarien abgrenzen.

vgl.²¹⁶ (Peterjohann, 2023)

Schritt 2: BSI-Grundschutz-Anforderungen bezieht sich auf die Umsetzung der BSI-Grundschutz-Anforderungen. Hierbei werden die Schutzziele den spezifischen Anforderungen zugeordnet, die im **BSI-Grundschutz-Baustein SYS** behandelt werden. Durch den Einsatz von Kreuzreferenztabellen wird ermittelt, welche Anforderungen besonders geeignet sind, um potenzielle Gefahren abzudecken. Zudem wird priorisiert, welche Schutzziele vorrangig den Anforderungen entsprechen, um eine effektive Eindämmung der Gefahren zu gewährleisten.

Diese systematische Vorgehensweise ermöglicht eine umfassende Analyse und Ausrichtung auf die individuellen Schutzbedarfe eines Unternehmens. Durch die klare Zuordnung zu BSI-Grundschutz-Anforderungen wird gewährleistet, dass die Sicherheitsmaßnahmen präzise auf die spezifischen Gefährdungen abgestimmt sind, um einen effizienten Schutz vor möglichen Schadensfällen zu gewährleisten. Die Verletzung dieser Schutzziele durch Angreifer kann Schwachstellen im IT-System offenbaren oder den Eindringen in ein IT-System ermöglichen. Die Gefahr einer realen Bedrohung der Unternehmenswerte, Datenspionage oder -manipulation ist hoch. Daher ist ein proaktives Risikomanagement erforderlich, um geeignete IT-Sicherheitsmaßnahmen zu ergreifen und Standards festzulegen.²¹⁷

Risikobewertung

Einschätzung des Schutzbedarfs wird laut dem BSI anhand des Risikos bewertet:

gering: Vorhandene oder geplanten Maßnahmen gewährleisten einen angemessenen Schutz gemäß dem Sicherheitskonzept.

mittel: Bereits umgesetzten oder geplanten Maßnahmen könnten möglicherweise nicht ausreichend sein und erfordern eine eingehendere Prüfung im Sicherheitskonzept.

hoch: Bestehenden oder geplanten Sicherheitsmaßnahmen bieten keinen ausreichenden Schutz vor der spezifischen Bedrohung und das Risiko ist mit hoher Wahrscheinlichkeit nicht akzeptabel.

sehr hoch: Implementierter oder geplanter Sicherheitsmaßnahmen besteht ein unzureichender Schutz vor der jeweiligen Gefährdung, sodass das Risiko mit sehr hoher Wahrscheinlichkeit nicht toleriert werden kann.

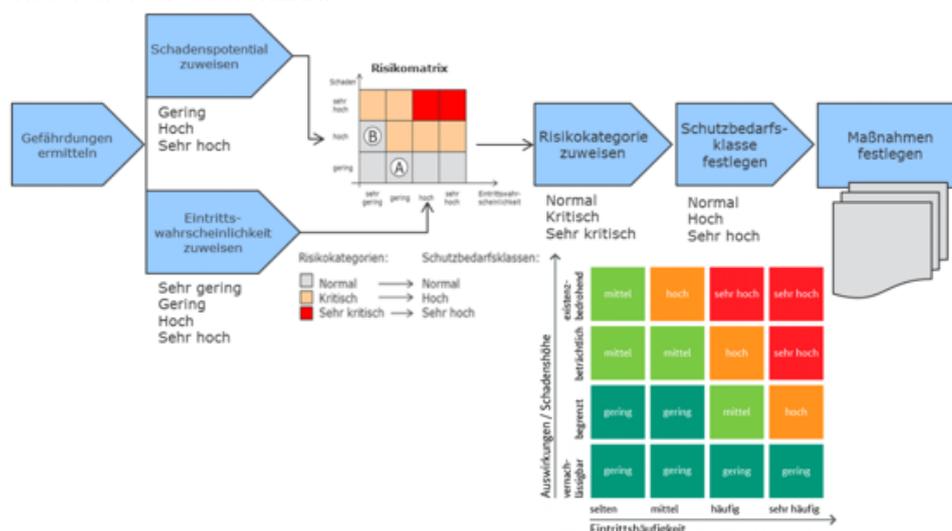


Abbildung 158: Risikomatrix mit Risikoeinstufung

Das BSI listet alle elementaren Gefährdungen auf, die die Schutzziele im Unternehmen beeinträchtigen können. Dazu gehören Datenverlust, Missbrauch von Berechtigungen, Diebstahl von Geräten und Datenträgern, Verlust von Geräten und Datenträgern, Offenlegung

vgl.²¹⁷ (Peterjohann, 2023)

schützenswerter Informationen, Verstoß gegen Gesetze oder Regelungen sowie Manipulation von Informationen.

Ein strukturierter Ansatz zur Festlegung von Schutzbedarfen und zur Umsetzung von BSI-Grundschutz-Anforderungen ist entscheidend. Unternehmen sollten präventive Maßnahmen einführen, um zukünftigen Schadensfällen besser entgegenzuwirken.²¹⁸

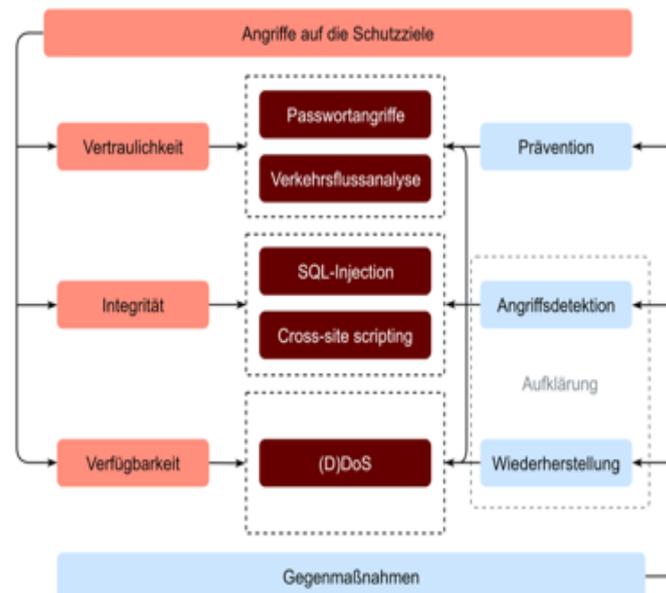


Abbildung 159: Zusammenhang zwischen Angriffen auf die Schutzziele und Gegenmaßnahmen

Schwachstellen nach BSI

Nachfolgend sind Schwachstellen aufgeführt, die nach Angaben des Bundesamts für Sicherheit in der Informationstechnik (BSI) besonders weit verbreitet sind:²¹⁹

Software-Schnittstellen: Programmierfehler stellen Angriffspunkte für Cyberkriminelle dar.

Design-Schnittstellen: Veralteter Code ermöglicht den unbefugten Zugriff auf Zugriffsrechte, Schnittstellen, Datenformate und Übertragungsprotokolle.

Konfigurationsschwachstellen: Schwächen bei der Implementierung von Software und IT-Systemen.

Menschliches Fehlverhalten und Schwächen bei Mitarbeitern: Durch Schulungen sollen Mitarbeiter auf menschliche Fehler im Zusammenhang mit Social Engineering, Phishing-Mails und gefälschten Links sensibilisiert und reduziert werden.

Elemente einer Cybersecurity-Strategie: Schutz vor Angriffen sowie Schutz von Geschäftsprozessen, geistigem Eigentum und anderen sensiblen Unternehmensdaten.

Kenntnis und Bewertung von Risiken: Compliance und Risikomanagement müssen durchgeführt werden, etwa durch das Monitoring des gesamten Unternehmensökosystems.

Incident Response und Business Continuity Management: Es müssen Strategien, Pläne, Maßnahmen und Prozesse entwickelt werden, um Schäden durch die Unterbrechung des IT-Betriebs zu minimieren. Notfallpläne und Backup-Systeme sollen dazu dienen, Daten zeitnah zu sichern und wiederherzustellen.

vgl.²¹⁸ (BSI, 2023)

vgl.²¹⁹ (Zillmann, Mario; Partner, Lünendonk & Hossenfelder, 2020)

Elemente einer Cyber-Security-Strategie

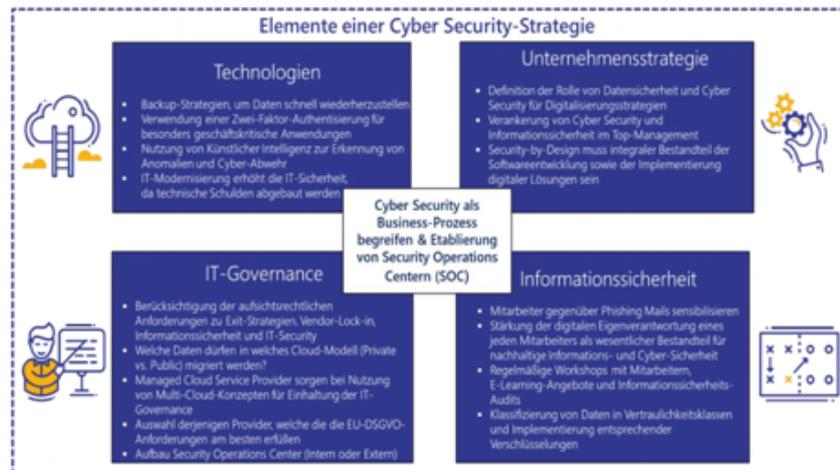


Abbildung 160: Elemente einer Cyber-Security-Strategie

Das reibungslose Zusammenspiel von Technologien, Unternehmensstrategie, IT-Governance und Informationssicherheit bildet die Grundlagen einer Sicherheitsstrategie, die als geschäftskritischer Prozess betrachtet wird. Die Implementierung eines SOC-System spielt hierbei eine zentrale Rolle.

Beispielhafte **Technologien**, die in diesem Kontext eingesetzt werden können, umfassen Backup-Strategien zur schnellen Wiederherstellung von Daten, die Anwendung einer Zwei-Faktor-Authentisierung für besonders geschäftskritische Anwendungen, die Nutzung von KI zur Erkennung von Anomalien und zur Cyberabwehr sowie die IT-Modernisierung, die die IT-Sicherheit durch den Abbau technischer Schulden erhöht.

Die **Unternehmensstrategie** legt die Bedeutung von Datensicherheit und Cybersecurity im Kontext der Digitalisierungsstrategie fest. Dabei erfolgt die Verankerung von Cybersecurity und Informationssicherheit auf höchster Managementebene. Des Weiteren sollte Security-by-Design ein integrierter Bestandteil sowohl der Softwareentwicklung als auch der Implementierung digitaler Lösungen sein.

Gerade bei der **Kritische Infrastrukturen (KRITIS)** wo eine grosse Sicherheitsbetrachtung gelegt wird auf Einrichtungen und Systeme, deren Ausfall oder Beeinträchtigung erhebliche Auswirkungen auf die Versorgungssicherheit, die öffentliche Sicherheit oder andere wichtige Funktionen der Gesellschaft haben könnte, muss der Schutz dieser kritischer Infrastrukturen entscheidend für die Aufrechterhaltung des täglichen Lebens und für die nationale Sicherheit besondere Sicherheitsvorkehrungen durchgeführt werden.

Die **IT-Governance** umfasst die Berücksichtigung verschiedener aufsichtsrechtlicher Anforderungen wie Exit-Strategien, Vendor-Lock-in, sowie Aspekte der Informationssicherheit und IT-Security. Hierzu zählt die Festlegung, welche Daten in welchem Cloud-Modell (Private vs. Public) migriert werden dürfen. Bei der Nutzung von Multi-Cloud-Konzepten gewährleisten Managed Cloud Service Provider die Einhaltung der IT-Governance. Die Auswahl dieser Anbieter erfolgte unter Berücksichtigung ihrer Fähigkeit, die Anforderungen der EU-DSGVO optimal zu erfüllen. Ein internes und externes **Security Operations Center (SOC)** trägt zur umfassenden Sicherheit bei.

Im Bereich der **Informationssicherheit** gilt es, Mitarbeiter gegenüber Phishing-Mails zu sensibilisieren und die digitale Eigenverantwortung jedes Einzelnen zu stärken. Dies sind entscheidende Elemente für eine nachhaltige Informations- und Cybersicherheit. Dazu gehören regelmäßige Workshops, E-Learning-Angebote und Informationssicherheits-Audits. Die Klassifizierung von Daten in vertrauliche Kategorien und die Implementierung entsprechender Verschlüsselungsmaßnahmen tragen ebenfalls dazu bei.²²⁰

vgl.²²⁰ (Hossenfelder, et al., 2021)

Top 10 Cyberangriffe 2023



Abbildung 161: Top 10 Cyberangriffe 2023

Bedrohungslage und Risiko bei Cyberangriffen laut BSI 2023

Laut dem Lagebericht des BMI²²¹ vom November 2023 als sehr hoch eingestuft wie nie zuvor. Dabei werden Schadprogramme-Varianten mit 332.000 neuen Varianten pro Tag ermittelt.

Die unterstehenden Grafiken geben einen Überblick über die IT-Sicherheitslage in Deutschland 2023.

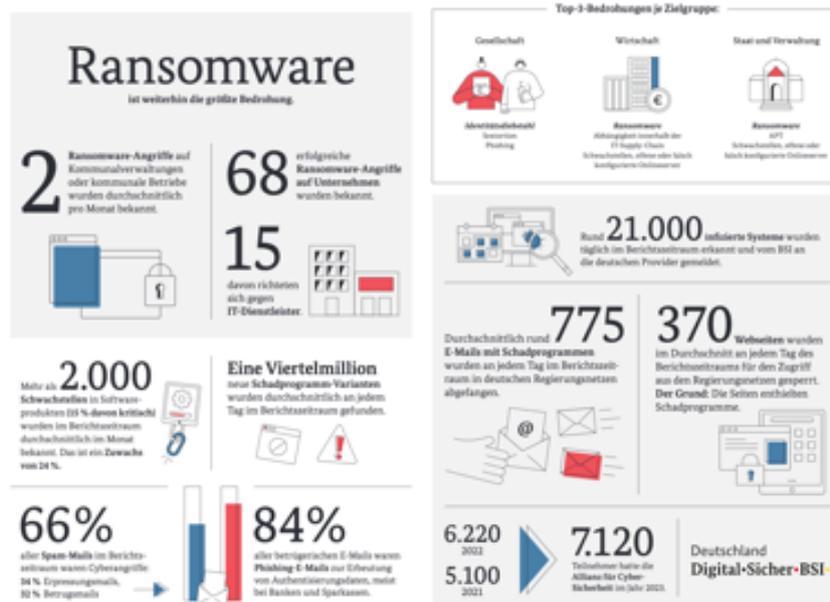


Abbildung 162: Die Lage der IT-Sicherheit in Deutschland 2023 im Überblick und Bedrohungsziele laut BSI

Die **Top 15 Cyber-Threats** treten am häufigsten auf, wie Malware, Web-basierende Angriffe, Phishing, Web-Applikationen-Angriffe, Spam, DDos, Identitätsdiebstahl, Datenverletzung, Insider-Bedrohung, Botnets, physische Manipulation, Schäden, Diebstahl und Verlust, Informationsleakage, Ransomware, Cyber-Spionage, Cyberjacking. Durch den Digitalisierungsprozess werden Schlüsseltechnologien mit Integration von KI und Deep Learning, Cloud, eID oder Smart Marketing oder modernen schnellen Netzwerken entwickelt, um kritische Infrastrukturen und die Sicherheit der Gesellschaft zu schützen.²²²

vgl.²²¹ (Janina Kröger)
vgl.²²² (Janina Kröger)

Typen von Angreifern

Gemäß dem BSI lassen sich **verschiedene Angreifergruppen im Cyber-Raum** identifizieren (Auszug).²²³

Cyber-Aktivisten (Hacktivisten) führen Cyberangriffe durch, um politische, gesellschaftliche, soziale, wirtschaftliche oder technische Themen bekannt zu machen sowie setzen Forderungen durch oder versuchen, Einfluss zu nehmen.

Cyber-Kriminelle verdienen illegal Geld in der IT und verursachen geringe Schäden sowie treten meist als Einzelpersonen oder in kleinen Gruppen mit begrenzter Professionalität auf.

Organisierte Cyber-Kriminalität verübt Identitätsdiebstahl von Bankdaten für Erpressung mit hoher Professionalität.

Konkurrenzausspähung/Industriespionage mit dem Ziel ifinanzieller Vorteil eines Unternehmenswettbewerbers oder privater Akteure sowie durch interne Daten- und Informationsauspähung entstehen Geldvorteile im weltweiten Wettbewerb.

Staatliche Nachrichtendienste nutzen staatlich gesteuerte wirtschaftliche Spionage für internationale Marktvorteile.

Staatliche Akteure im Cyber-War führen Cyberangriffe im militärischen Sektor des Cyber-Raums durch, z.B. auf Domänen wie Land, See, Luft und Weltraum.

Cyber-Terroristen nutzen verschiedene Ziele, um ihre Ideologie zu verbreiten und ihren Einfluss auszuweiten.

Hobbyisten/Skript-Kiddies führen Cyberangriffe aus Interesse und zum Testen ihrer Fähigkeiten durch sowie keine finanziellen Interessen, Angriffe variieren stark und sind oft von der Absicherung abhängig.

Innentäter (externe Dienstleister) agieren von außerhalb und haben bereits Zugang zu internen Ressourcen im Unternehmen und analysieren Schutzmaßnahmen und Schwachstellen über einen längeren Zeitraum.

IT-Sicherheitsforscher suchen primär nach akademischen Sicherheitslücken und Cyberangriffen sowie unkoordinierte Veröffentlichung von "Full Disclosure" kann von anderen Angreifern für reale Attacks genutzt werden.

Beispiel Ransomware-Cyber-Angriffe und Schutzmaßnahmen Ransomware-Cyber-Angriffe-Phasen

Ein **typischer Angriffsverlauf** beispielsweise die **Ransomware Alphv**, auch bekannt als BlackCat, wird von einer anderen Gruppierung als der Ransomware LockBit 3.0 eingesetzt. Der Ransomware-Angriff durchläuft verschiedene Phasen.²²⁴

Erstinfektion (Angriffsphase 1) beginnt der Angriff oft mit einer bösartigen E-Mail, der Kompromittierung eines Fernzugriffszugangs wie Remote Desktop Protocol (RDP) oder der Ausnutzung von Software-Schwachstellen. Diese initiale Infektion bildet den Ausgangspunkt für die weiteren Schritte des Angreifers.

Rechteerweiterung und Ausbreitung (Angriffsphasen 2 und 3) haben nach dem Einbruch der Angreifer nur die Zugriffsrechte des kompromittierten Accounts. Zusätzlich installiert er oft weitere Schadsoftware, um Zugriffsrechte zu erweitern. Der Angreifer breitet sich automatisiert im Netzwerk aus, bis er zentrale Rechteeverwaltungskomponenten wie das Active Directory übernimmt.

Datenabfluss (Angriffsphase 4) stehlen die Angreifer Daten (Datenexfiltration), drohen mit Veröffentlichung und erpressen das Opfer zu Lösegeld- oder Schweigegeldzahlungen.

Verschlüsselung (Angriffsphase 5) werden Daten auf vielen Systemen verschlüsselt, einschließlich Backup-Systemen. Die Angreifer hinterlassen Nachrichten für Lösegeldverhandlungen. Einige Gruppen erpressen direkt mit gestohlenen Daten, ohne sie zu verschlüsseln.

vgl.²²³ (BSI, 2023)

vgl.²²⁴ (COO)

Incident Response (Angriffsphase 6) stehen Betroffene vor der Herausforderung, Systeme und Daten wiederherzustellen. Je nach Ausmaß wird ein IT-Sicherheitsdienstleister hinzugezogen, um den Vorfall zu bewältigen und die Stakeholder zu informieren.

Beispiel Ransomware-Angriffe-Schutzmaßnahmen

Ein **Ransomware-Angriff** durchläuft verschiedene Phasen, für die gezielte Schutzmaßnahmen erforderlich sind. Hier sind die Maßnahmen für jede Angriffsphase.²²⁵

Angriffsphase 1 – Einbruch:

Gegenmaßnahme Malware-Spam:

- E-Mails sollten als "Nur-Text" angezeigt werden, um mögliche schadhafte Inhalte zu blockieren.
- Sensibilisierung der Mitarbeiter bezüglich E-Mail-Risiken ist entscheidend.

Gegenmaßnahme Schwachstellen:

- Sofortiges Einspielen von Sicherheitsupdates, insbesondere für kritische Schwachstellen.

Gegenmaßnahme Remote-Zugang:

- Einsatz von Multifaktor-Authentifizierung für Remote-Zugänge.

Angriffsphase 2 – Rechteerweiterung:

Gegenmaßnahme:

- Einschränkung der Verwendung privilegierter Accounts für administrative Tätigkeiten.
- Multifaktor-Authentifizierung für privilegierte Konten.

Angriffsphase 3 – Ausbreitung:

Gegenmaßnahme:

- Segmentierung des Netzwerks, um die Ausbreitung von Ransomware zu begrenzen.
- Sicherer Einsatz von Administrator-Accounts.

Angriffsphase 4 – Datenabfluss:

Gegenmaßnahme:

- Anomalie-Detektion im Netzwerk zur Früherkennung unerwünschter Datenabflüsse.

Angriffsphase 5 – Verschlüsselung:

Gegenmaßnahme:

- Regelmäßige Offline-Backups zur sofortigen Datenwiederherstellung nach einer Verschlüsselung.
- Planung und regelmäßiges Testen der Backup-Wiederherstellungsprozesse.

Angriffsphase 6 – Incident Response:

Gegenmaßnahme:

- Vorhandensein eines Notfallplans für den Worst-Case eines flächendeckenden Netzwerkangriffs.
- Regelmäßige Schulung und Übung der Prozesse zur Reaktion und Wiederherstellung.

Diese präventiven Maßnahmen minimieren das Risiko und unterstützen die schnelle Reaktion auf Ransomware-Angriffe.²²⁶

vgl.²²⁵ (BSI, 2023)
vgl.²²⁶ (BSI, 2023)



Abbildung 163: Threat – Asset – Vulnerability – Risk - Zusammenhänge

Der Zusammenhang zwischen Bedrohung (Threat), Vermögenswert (Asset) und Schwachstelle (Vulnerability) ist fundamental für das Verständnis und die Bewertung von Risiken in der Informationssicherheit. **A + T + V = Risk**

Bedrohung (Threat-T) ist eine Bedrohung bezieht sich auf potenzielle Gefahren oder schädliche Ereignisse, die die Sicherheit eines Systems oder Vermögenswerts beeinträchtigen könnten. Dies könnten beispielsweise Cyberangriffe, Naturkatastrophen, menschliche Fehler oder böswillige Handlungen sein.

Vermögenswert (Asset-A) ist ein Vermögenswert ist jede Ressource, Information oder Entität, die einen Wert für eine Organisation hat. Dies können Daten, Hardware, Software, Mitarbeiter, Reputation oder physische Räumlichkeiten sein. Der Schutz von Vermögenswerten ist entscheidend für den Geschäftsbetrieb und den Unternehmenserfolg.

Schwachstelle (Vulnerability-V) ist eine Schwachstelle ist eine Schwäche oder ein Fehler in einem System, einer Anwendung oder einer organisatorischen Praxis, die von Bedrohungen ausgenutzt werden kann, um einen Vermögenswert zu gefährden. Schwachstellen können beispielsweise unsichere Konfigurationen, fehlende Patches oder Schwächen in Prozessen sein.

Risiko (Risk-R) ist die Wahrscheinlichkeit, dass eine Bedrohung eine Schwachstelle ausnutzt und einen Schaden an einem Vermögenswert verursacht. Es wird oft als Produkt aus der Eintrittswahrscheinlichkeit einer Bedrohung, der Ausnutzbarkeit einer Schwachstelle und dem potenziellen Schaden für den Vermögenswert definiert. Risikobewertungen helfen, Prioritäten zu setzen und geeignete Sicherheitsmaßnahmen zu ergreifen.

Der Zusammenhang zwischen Bedrohung, Vermögenswert und Schwachstelle verdeutlicht, dass das Risiko in der Informationssicherheit nicht nur von externen Gefahren abhängt, sondern auch von internen Schwächen in den Systemen und Prozessen. Das Identifizieren, Bewerten und Minimieren dieser Risiken ist entscheidend, um die Sicherheit von Vermögenswerten zu gewährleisten und Geschäftsunterbrechungen zu verhindern.

Anhang IV Abkürzungsverzeichnis

5G	5-Generation Standard
AAE	Adversarial Autoencoder
ACSC	Australian Cyber Security Centre
AES	Advanced Encryption Standard
AFGBV	Autonome-Fahrzeuge-Genehmigungs-und-Betriebs-Verordnung
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AIOps	Artificial Intelligence for IT Operations
ANI	Artificial Narrow Intelligence
APM	Application Performance Monitoring
API	Application Programming Interface
APM	Application Performance Monitoring
ASI	Artificial SuperintelligenceIntelligence
AQL	Ariel Query Language
AR	Augmented Reality
ASIO	Australian Security Intelligence Organisation
AWS	Amazon Web Services
BCS	Business Continuity System
BMBF	Bundesministerium für Bildung und Forschung
BMI	Bundesministerium des Inneren
BMU	Best Matching Unit
BMWi	Bundesministerium für Wirtschaft und Energie
BMVg	Bundesministerium der Verteidigung
BSI	Bundesamt für Sicherheit in der Informationstechnik
BYOD	Bring Your Own Device
CAPTCHA	Completely Automated Public Turing Test to tell Computers and Humans Apart
CASB	Cloud Access Security Broker
CD	Compact Disc
CD	Continuous Delivery/Continuous Deployment
CE	Conformité Européenne
CERT	Computer Emergency Response Team
CI	Continuous Integration
CIS	Center for Internet Security
CMDB	Configuration Management Database
CNN	Convolutional Neural Network
COBIT 5	Control Objectives for Information and Related Technologies 5
CPAI	Global Partnership on Artificial Intelligence
CPU	Central Processing Unit
CRISP-DM	Cross-Industry Standard Process for Data Mining
CSIRP	Computer Security Incident Response Plan
CSIRT	Computer Security Incident Response Team
CSV	Comma-separated values
CTEM	Continuous Threat Exposure Management
DBN	Deep Belief Network
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DDoS	Distributed Denial of Service
DER	Endpoint Detection and Response
DEM	Digital Experience Monitoring
DevOps	Development Operations
DISA	Direct Inward System Access
DL	Deep Learning
DLL	Dynamic Link Library

DLR	Deutsche Zentrum für Luft- und Raumfahrt
DLOP	Disturbed Line Of Sight
DNS	Domain Name System
DSDL	Data Science and Deep Learning
DSGVO	Datenschutz-Grundverordnung
EDR	Endpoint Detection and Response
eID	electronic Identity
EU	European Union
EU-DSGVO	European Union-Datenschutz-Grundverordnung
EUE	End User Equipment
EXE	executable
FFN	Feedforward-Neuronale Networks
FI CODE	Forschungsinstitut Cyber Defence
FIDO2	Fast Identity Dentity Online2
FPR	False Positive Rate
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GIGO	Garbage out-Prinzip
GPAI	General Purpose Artificial Intelligence
GUI	Graphical User Interface
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
HIPAA	Health Insurance Portability and Accountability Act
IAM	Identity and Access Management
IAIS	Fraunhofer-Institut für Intelligente Analyse-und Informationssysteme
IBM	International Business Machines
IBM Db2	International Business Machines Datenbankmanagementsystem2
ICT	Information and Communication Technology
IDPS	Intrusion-Detection- und Intrusion-Prevention-System
IDS	Intrusion Detection Scan
IEC	International Electrotechnical Commission
IIoT	Industrial Internet of Things
IMS	International Monetary Fund
IoT	Internet der Dinge
IP	Internet Protocol
IPS	Intrusion Prevention System
IQWIG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IR	Incident Response
ISACA	Information Systems Audit and Control Association
ISF	Information Security Forum
ISMS	Information Security Management System
ISO	International Organization for Standardization
IT	Internetwork Technology
ITIL	Information Technology Infrastructure Library
ITOA	Internetwork Technology Operations Analytics
ITSM	Internetwork Technology Service Management
JRC	Joint Research Centre
JSON	JavaScript Object Notation
KI	Künstlicher Intelligenz (Artificial intelligence)
KI-VO	Künstlicher Intelligenz-Verordnung
KMU	kleine und mittlere Unternehmen
KNN	K-Nearest Neighbors
kNN	K-Nearest Neighbors
KPIs	Key Performance Indicators
KRITIS	Critical Infrastructures

LAN	Local Area Network
LIME	Local Interpretable Model-Agnostic Explanations
LOLBins	Living Off The Land Binaries
LSM	Liquid State Machine
LSQ	Least Squares
LSTM	Long Short-Term Memory
MASA	Mesh-App- und Service-Architekturen
MDR	Managed Detection and Response
MDR	Medizinprodukte-Verordnung
MFA	Multi-Faktor-Authentifizierung
ML	Maschine Learning
MLOP	Maschine Learning Operation
MLP	Multilayer Perceptron (MLP) Neural Network
MLTK	Machine Learning Toolkit
MNN	Modular Neural Network
MMS	Multimedia Messaging Service
MSI	Microsoft Software Installer
MSP	Managed Service Provider
MTTD	Mean Time To Detect
MTTR	Mean-Time-To- Reaktion
MySQL	My Structured Query Language
NDR	Network Detection and Response
NIS2	Network and Information Security Directive 2
NIST	National Institute of Standards and Technology
NKCS	National Coordination Centre for Cybersecurity
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
NOSQL	not only SQL
NSA	National Security Agency
NTA	Network Traffic Analysis
ODBC	Open Database Connectivity
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
OGC	Office of Government Commerce
OPC UA	Open Platform Communications Unified Architecture
OS	Operating System
OT	Operational Technology
OTP	One-Time Passwords
PCI	Pre-Compartmented Information
PCI DSS	Payment Card Industry Data Security Standard
PDF	Portable Document Format
PKI	Public Key Infrastruktur
PoC	Proof of Concept
QML	Quantum-Machine-Learning
RBF	Radial Basis Function Neural Network

RBM	Restricted Boltzmann Machines-Neural Network
RDP	Remote Desktop Protocol
ResNet	Residual Neural Network
RNN	Recurrent Neural Network
OPERLOG	Operator Log
OTP	One-Time Password
SaaS	Software as a Service
SAML	Security Assertion Markup Language
SAP	Systeme, Anwendungen, Produkte
SEM	Security Event Management
SEO	Suchmaschinenoptimierung
SIM	Security Information Management
SIEM	Security Information and Event Management
SHAP	SHapley Additive exPlanation
SMF	System Management Facility
SMS	Short Message Service
SNN	Spiking Neural Network
SOAR	Security Orchestration, Automation and Response
SOC	Security Operation Center
SOCaaS	Security Operations Center as a Service
SoGP	Security Technical Implementation Guides
SOM	Self-organizing Maps
SPL	Software Product Line
SSO	Single Sign-On
STIG	Security Technical Implementation Guide
SVM	Support Vector Machine
SYS	Systeme (BSI-Baustein)
SYSLOG	System Logging Protocol
TIS	Threat Intelligence Service
TRISM	Trust, Risk and Security Management
TPR	True Positive Rate
UEBA	User- und Entity-Behavior-Analytics
USB	Universal Serial Bus
UX	User-Experience-Design
VAE	Variational Autoencoder
VMS	Vulnerability Management
VPN	Virtual Private Network
VR	Virtual Reality
WLAN	Wireless Local Area Network
XAI	Explainable Artificial Intelligence
XDR	Extended Detection and Response

Anhang V Abbildungsverzeichnis

Abbildung 1: Kernbestandteile der Cyber-Security

Zillmann, Mario; Partner, Lünenendok & Hossenfelder GmbH. Cyber Security. - Die digitale Transformation sicher gestalten.

<https://s3.eu-central-1.amazonaws.com/cdn.a3bau.at/public/2021-02/Whitepaper-Luenendok-ArvatoSystems-CyberSecurity.pdf>, 2020, S.15

Abbildung 2: SecurityInformation and Event Management (SIEM)

WOLF, ARCTIC. THE MARKET LEADER IN SECURITY OPERATIONS.

<https://arcticwolf.com/resource/aw/siem-a-comprehensive-guide>, 2023

Abbildung 3: SIEM- Funktionen und Service

Mohan, Remya. What Is Security Information and Event Management (SIEM)? Definition, Architecture, Operational Process, and Best Practices. SIEM manages vulnerabilities by providing next-generation detection, analytics, and response.

<https://www.spiceworks.com/it-security/vulnerability-management/articles/what-is-siem/>, 2022

Kidd, Chrissy. SIEM: Security Information & Event Management Explained.

https://www.splunk.com/en_us/blog/learn/siem-security-information-event-management.html, 2023

Abbildung 4: SIEMs Architektur

Mohan, Remya. What Is Security Information and Event Management (SIEM)? Definition, Architecture, Operational Process, and Best Practices. SIEM manages vulnerabilities by providing next-generation detection, analytics, and response.

<https://www.spiceworks.com/it-security/vulnerability-management/articles/what-is-siem/>, 2022

Abbildung 5: Darstellung einer SIEM-Architektur

Brombacher, Jürgen. Aufbau eines SIEM- und SOC-Systems.

https://www.matrix.ag/blog/aufbau-eines-siem-und-soc-systems?gclid=EAlaIqobChMIkKj6y-T9ggMVWJaDBx1wCAK1EAMYASAAEgJ29vD_BwE

Abbildung 6: Systemkomponenten eines SIEM

KOGIT. Security Information & Event Management.

<https://www.kogit.de/services/security-information-event-management/>

Abbildung 7: Workflow eines SIEM

KOGIT. Security Information & Event Management.

<https://www.kogit.de/services/security-information-event-management/>

Abbildung 8: SIEM Maturity Hierachy

Miuccio, Tony. SIEM Maturity Hierarchy.

<https://blacktowersec.com/siem-maturity-hierarchy/>

Abbildung 9: Team SIEM

Grüner, Tina. Was ist eigentlich dieses „Managed SIEM“?

<https://blog.to.com/managed-siem/>, 2016

Abbildung 10: SIEM-Prozesse

Mohan, Remya. What Is Security Information and Event Management (SIEM)? Definition, Architecture, Operational Process, and Best Practices. SIEM manages vulnerabilities by providing next-generation detection, analytics, and response.

<https://www.spiceworks.com/it-security/vulnerability-management/articles/what-is-siem/>, 2022

Abbildung 11: Roadmap Erstellung SIEM Grobkonzept nach CBT Training & Consulting GmbH

Brinz, Dipl.-Inf. Christian. CBT Training & Consulting GmbH.

Schulungsunterlagen Kursunterlagen_SIEM nach ISO27001-05-2023, 2023, S.107

Abbildung 12: Kickoff- Programmplan

Brombacher, Jürgen. Aufbau eines SIEM- und SOC-Systems.

https://www.matrix.ag/blog/aufbau-eines-siem-und-soc-systems?gclid=EAlaIQobChMItKj6y-T9ggMVWJaDBx1wCAK1EAMYASAAEgJ29vD_BwE

Abbildung 13: am häufigsten delegierten SOC-Anwendungsfälle

Infopulse. SOC-Einführung: Drei Szenarien für eine bessere Sicherheitslage.

<https://infopulsemarketing.blob.core.windows.net/ebooks-reports/de-soc-einfuehrung-drei-Szenarien-fuer-eine-bessere-sicherheitslage.pdf>, S.27

Abbildung 14: Projektschritte für die Einführung von SIEMs nach CBT Training & Consulting GmbH

Brinz, Dipl.-Inf. Christian. CBT Training & Consulting GmbH.

Schulungsunterlagen Kursunterlagen_SIEM nach ISO27001-05-2023, 2023, S.105

Abbildung 15: Angriffserkennungen von SIEMs nach CBT-Training & Consulting GmbH

Brinz, Dipl.-Inf. Christian. CBT Training & Consulting GmbH.

Schulungsunterlagen Kursunterlagen_SIEM nach ISO27001-05-2023, 2023, S.110

Abbildung 16: Übersicht SIEM & SOAR

Diener, Alexandra. Security Operation Center Lab.

<https://eprints.ost.ch/id/eprint/976/1/FS%202021-BA-EP-Diener-Security%20Operation%20Center%20Lab.pdf>, 2021, S.10

Abbildung 17: Gartner Bewertung von Tools 2023

Gregg, Siegfried; Bangera, Mrudula, Crossley, Matt; Byrne, Padraig. Magic Quadrant for Application Performance Monitoring and Observability.

<https://www.gartner.com/doc/reprints?id=1-2EDYKN6L&ct=230706&st=sb>, 2023

Abbildung 18: IBM QRadar

INFOGUARD AG. IBM QRADAR – SECURITY INFORMATION & EVENT MANAGEMENT PLATFORM (SIEM).

<https://www.infoguard.ch/de/partner/ibm-qradar-security-information-event-management-siem>

Abbildung 19: IBM QRadar Types of content extensions

IBM. QRadar content extensions.

<https://www.ibm.com/docs/en/qsip/7.4?topic=qradar-content-extensions>, 2024

Abbildung 20: AQL query flow

IBM. AQL Query structure.

<https://www.ibm.com/docs/en/qsip/7.4?topic=aql-query-structure>, 2023

Abbildung 21: Simple AQL queries

IBM. Sample AQL queries.

<https://www.ibm.com/docs/en/qsip/7.4?topic=structure-sample-aql-queries>, 2023

Abbildung 22: QRadar Architektur

IBM. QRadar architecture overview.

<https://www.ibm.com/docs/en/qsip/7.4?topic=deployment-qradar-architecture-overview>, 2023

Abbildung 23: Problem insights overview

IBM. Problem insights overview.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-problem-insights>, 2024

Abbildung 24: Metric-based machine learning on z/OS

IBM. Metric-based machine learning overview.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-metric-based-machine-learning>, 2024

Abbildung 25: Log-based machine learning on Linux

IBM. Metric-based machine learning overview.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-metric-based-machine-learning>, 2024

Abbildung 26: IBM Security QRadar SIEM Demo

IBM. IBM Security QRadar SIEM demo.

https://mediacenter.ibm.com/media/IBM+Security+QRadar+SIEM+demo/1_yqg5lInj

Abbildung 27: Logpoint Plattformen

Logpoint. Bedrohungen schnell erkennen und entschärfen.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 28: Erkennung kompromittierter Benutzer-Anmeldedaten

Logpoint. Development That Works for You. (Grafik angepasst)

<https://www.logpoint.com/en/deployment/>

Abbildung 29: Nachverfolgung von Systemänderungen

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 30: Erkennung von ungewöhnlichem Verhalten bei privilegierten Konten

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 31: Sicherheit für cloudbasierte Anwendungen

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 32: Erkennung von Phishing-Angriffen

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 33: Überwachung von Auslastung und Verfügbarkeit

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 34: Logdaten-Management

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 35: SIEM für GDPR, HIPAA oder PCI-Compliance

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 36: Suche nach Bedrohungen (Threat Hunting)

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 37: SIEM für die Automatisierung

Logpoint. Die Top 10 SIEM-Use-Cases für Ihre Implementierung.

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Abbildung 38: Logpoint SIEM

Logpoint. Mit einem einfachen Add-on von SIEM zur ganzheitlichen Cyber Defense.

https://www.logpoint.com/de/?utm_source=google&utm_campaign=DE_DACH_Brand&utm_medium=cpc&utm_content=DE_DACH_Brand_PM&utm_term=logpoint&gad_source=1&gclid=CjwKCAiAnL-sBhBnEiwAJRGigtAKOAAm7osKL30NyGe_23z1LnJGFsxTHi3UVSti-EqB7IjkV35JDRoC_0MQAvD_BwE

Abbildung 39: Logpoint SIEM Platform Solution

Logpoint. SIEM Buyer's Guide 2023.

<https://www.logpoint.com/wp-content/uploads/2023/05/siem-buyers-guide-2023.pdf>, 2023, S.2

Abbildung 40: wichtige Anwendungsfälle

Logpoint. SIEM Buyer's Guide 2023.

<https://www.logpoint.com/wp-content/uploads/2023/05/siem-buyers-guide-2023.pdf>, 2023, S.4, S.5

Abbildung 41: SIEM vs. UEBA

Logpoint. Was ist User and Entity Behavior Analytics? Ein umfassender Leitfaden zu UEBA, der Funktionsweise und den Vorteilen.

<https://www.logpoint.com/de/blog/what-is-ueba-a-complete-guide-to-ueba/>, 2020

Abbildung 42: UEB- Erkennung

Logpoint. Was ist User and Entity Behavior Analytics? Ein umfassender Leitfaden zu UEBA, der Funktionsweise und den Vorteilen.

<https://www.logpoint.com/de/blog/what-is-ueba-a-complete-guide-to-ueba/>, 2020

Abbildung 43: Best Practices for UEBA

Logpoint. Was ist User and Entity Behavior Analytics? Ein umfassender Leitfaden zu UEBA, der Funktionsweise und den Vorteilen.

<https://www.logpoint.com/de/blog/what-is-ueba-a-complete-guide-to-ueba/>, 2020

Abbildung 44: DarkGate Infektionskette

Bogati, Anish. Logpoint. DarkGate infection chain.

<https://www.logpoint.com/en/blog/inside-darkgate/>, 2024

Abbildung 45: Logpoint AgentX Isolate-Unisolate Host
Bogati, Anish. Logpoint. DarkGate infection chain.
<https://www.logpoint.com/en/blog/inside-darkgate/>, 2024

Abbildung 46 Logpiont SIEM Demo
Logpoint. SIEM reduziert das Cyber-Risiko mit leistungsstarken Datenanalysen.
<https://www.logpoint.com/de/produkt/siem/>

Abbildung 47: LogRhythm
LogRhythm. Your Trusted Security Partner.
<https://logrhythm.com/>

Abbildung 48: Evolution of SIEM-Software
LogRhythm. Security Information and Event Management (SIEM) Solutions.
<https://logrhythm.com/solutions/security/siem/>

Abbildung 49: LogRhythm Architektur
Ashwani, K.. What is LogRhythm and use cases of LogRhythm?
<https://www.devopsschool.com/blog/what-is-logrhythm-and-use-cases-of-logrhythm/>, 2023

Abbildung 50: LogRhythm SIEM Demo
LogRhythm. Demo: Gain Visibility and Threat Detection Across Hybrid Environment.
<https://logrhythm.com/securing-hybrid-environments/>

Abbildung 51: SolarWinds SIEM
SolarWinds. Observability done right. Finally.
<https://www.solarwinds.com/de/solarwinds-platform#>

Abbildung 52: künstliche Intelligenz für IT-Abläufe bei SolarWinds
SolarWinds. SolarWinds AIOps – integrierte Intelligence.
Intelligentere Lösungen aufbauend auf zwei Jahrzehnten Erfahrung.
<https://www.solarwinds.com/de/hybrid-cloud-observability/use-cases/aiops>

Abbildung 53: SolarWinds SIEM Demo
SolarWinds. Demo.
<https://sem.demo.solarwinds.com/webui/events/137816-2>

Abbildung 54: ManageEngine
ManageEngine. Digital headquarters for advanced data loss prevention.
<https://pdf.indiamart.com/impdf/2852588612148/MY-8100832/manageengine-dlp-plus.pdf>

ManageEngine. A single pane of glass for complete Endpoint Management and Security.
<https://download.manageengine.com/products/desktop-central/desktop-administration-overview.pdf>, 2024, S.1

Abbildung 55: Analyse des Benutzer- und Entitätsverhaltens mithilfe von KI und Prozessflussdiagramm für die Analyse von Benutzerentitäten und -verhalten
ManageEngine. Enterprise AIOps. How ManageEngine refines IT processes with artificial intelligence.
https://download.manageengine.com/academy/aops_e-book.pdf, S.27, S.28

Abbildung 56: Prozessflussdiagramm für die Vorhersage von Ausfällen
ManageEngine. Enterprise AIOps. How ManageEngine refines IT processes with artificial intelligence.
https://download.manageengine.com/academy/aops_e-book.pdf, S.23

Abbildung 57: MangeLogs - Audit - Secure – Be Compliant

NSITechnology. EventLogAnalyzer.

<https://www.nsitech.africa/ManageEngine/EventlogAnalyzer/eventlog-analyzer>

Abbildung 58: ManageEngine Demo

ManageEngine. OpManager Online Demo.

<https://demo.opmanager.com/>

Abbildung 59: Splunk Observability Cloud Schema

Splunk. Die Splunk-Plattform schafft End-to-End-Transparenz vom Edge bis in die Cloud.

https://www.splunk.com/de_de/products/splunk-enterprise.html

Abbildung 60: Splunk Observability

Splunk. Splunk Observability.

https://www.splunk.com/de_de/products/observability.html?301=/de_de/it-operations.html

Abbildung 61: Quellen von Spunk

Splunk. A Beginner's Guide to Observability. Cutting through the complexity to learn what your systems, services and apps are really doing.

<https://www.splunk.com/pdfs/ebooks/beginners-guide-to-observability.pdf>, 2021, S.8, S.9

Abbildung 62: Risikobewertung bei MLTK- oder Out-of-the-Box-Anwendungsfall

Splunk. Splunk AI: Use-Case-Leitfaden für Einsteiger - Künstliche Intelligenz von Splunk für Observability.

https://www.splunk.com/de_de/pdfs/gated/ebooks/splunk-machine-learning-for-observability-use-case-guide.pdf, 2023, S.7

Abbildung 63: ML-basierten Analysen in Splunk

Splunk. Splunk AI: Use-Case-Leitfaden für Einsteiger - Künstliche Intelligenz von Splunk für Observability.

https://www.splunk.com/de_de/pdfs/gated/ebooks/splunk-machine-learning-for-observability-use-case-guide.pdf, 2023, S.9

Abbildung 64: Vorhersage von Datenausfällen in Splunk

Splunk. Splunk AI: Use-Case-Leitfaden für Einsteiger - Künstliche Intelligenz von Splunk für Observability.

https://www.splunk.com/de_de/pdfs/gated/ebooks/splunk-machine-learning-for-observability-use-case-guide.pdf, 2023, S.18

Abbildung 65: Splunk Demo

Splunk. TURN DATA INTO DOING Splunk Observability Cloud Free Trial.

https://www.splunk.com/en_us/download/o11y-cloud-free-trial.html

Abbildung 66: drei wichtigsten Arten von KI

SAP. Was ist künstliche Intelligenz?

<https://www.sap.com/austria/products/artificial-intelligence/what-is-artificial-intelligence.html>

Abbildung 67: KI-Technologien

SAP. Was ist künstliche Intelligenz?

<https://www.sap.com/austria/products/artificial-intelligence/what-is-artificial-intelligence.html>

Abbildung 68: Leistungsbestandteile der Künstlichen Intelligenz

Management Circle. Das Grosse 1x1 der Künstlichen Intelligenz. Alles. Was Sie über KI, ChatGPT & Co. Wissen müssen.

<https://go.managementcircle.de/ki-1x1>, S.4, (Teilansicht)

Abbildung 69: Arten von Machine Learning Algorithmen

datasolut. Wuttke, Vinzent. Machine Learning: Definition, Algorithmen, Methoden und Beispiele.

<https://datasolut.com/was-ist-machine-learning/#unueberwachtetes-lernen>, 2024

Abbildung 70: Unsupervised Learning (Unüberwachtes Lernen) ist eine Art von Machine Learning, die eigenständig Muster und Zusammenhänge in den Daten findet

datasolut. Wuttke, Vinzent. Machine Learning: Definition, Algorithmen, Methoden und Beispiele.

<https://datasolut.com/was-ist-machine-learning/#unueberwachtetes-lernen>, 2024

Abbildung 71: Überwachtes maschinelles Lernen trainiert Muster und Zusammenhänge anhand von Daten mit einer Zielvariable

datasolut. Wuttke, Vinzent. Machine Learning: Definition, Algorithmen, Methoden und Beispiele.

<https://datasolut.com/was-ist-machine-learning/#unueberwachtetes-lernen>, 2024

Abbildung 72: Semi-überwachten Lernen

Medium. What is Semi-Supervised Learning? A Guide for Beginners.

<https://medium.com/@datasciencewizards/what-is-semi-supervised-learning-a-guide-for-beginners-a7452a597b8c>, 2023

Abbildung 73: einfaches Beispiel von verstärkendem Lernen durch Belohnungen

datasolut. Wuttke, Vinzent. Machine Learning: Definition, Algorithmen, Methoden und Beispiele.

<https://datasolut.com/was-ist-machine-learning/#unueberwachtetes-lernen>, 2024

Abbildung 74: Maschinelles Lernen im Überblick: Anwendungsbeispiele nach Arten

datasolut. Wuttke, Vinzent. Machine Learning: Definition, Algorithmen, Methoden und Beispiele.

<https://datasolut.com/was-ist-machine-learning/#unueberwachtetes-lernen>, 2024

Abbildung 75: neuronales Netzwerk

datasolut. Wuttke, Vinzent. Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion.

<https://datasolut.com/neuronale-netzwerke-einfuehrung/>

Abbildung 76: Einordnung neuronale Netz-Arten

Fischer, Prof. Dr. Jörn. Grundlagen Neuronale Netze.

https://services.informatik.hs-mannheim.de/~fischer/lectures/GNN_Files/GNN.pdf, S15

Abbildung 77: einfache und Multilayer neuronale Perceptron

CLOUDFLARE. Was ist ein neuronales Netzwerk?

<https://www.cloudflare.com/de-de/learning/ai/what-is-neural-network/>

Abbildung 78: Netzwerkdiagramm eines Feedforward-Netzes

Wallner, Anna. Neuronale Netze.

https://www.mathematik.uni-ulm.de/stochastik/lehre/ss07/seminar_sl/ausarbeitung_wallner.pdf, 2007, S.3

Abbildung 79: Faltung in Convolutional Neural Networks

datasolut. Wuttke, Vinzent. Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion.

<https://datasolut.com/neuronale-netzwerke-einfuehrung/>, 2024

Abbildung 80: Aufbau eines Recurrent Neural Networks und Long Short-Term Memory Units
Dataaspirant, Polamuri, Sharmila. LSTM: Introduction to long short term memory.
<https://dataaspirant.com/lstm-long-short-term-memory/#>

Abbildung 81: Modulare neuronale Netzwerke (MNNs)
CLOUDFLARE. Was ist ein neuronales Netzwerk?
<https://www.cloudflare.com/de-de/learning/ai/what-is-neural-network/>

Abbildung 82: Radialen Basisfunktionen-Neuronale Netzwerke
CLOUDFLARE. Was ist ein neuronales Netzwerk?
<https://www.cloudflare.com/de-de/learning/ai/what-is-neural-network/>

Abbildung 83: Liquid State Machine-Neuronale Netzwerke
CLOUDFLARE. Was ist ein neuronales Netzwerk?
<https://www.cloudflare.com/de-de/learning/ai/what-is-neural-network/>

Abbildung 84: Residuale-Neuronale Netzwerke
CLOUDFLARE. Was ist ein neuronales Netzwerk?
<https://www.cloudflare.com/de-de/learning/ai/what-is-neural-network/>

Abbildung 85: Generative Adversarial Networks
clickworker. Generative Adversarial Networks (GANs).
<https://www.clickworker.de/ki-glossar/generative-adversarial-networks/>

Abbildung 86: Self Organizing Maps
Geeksforgeek. Selbstorganisierende Karten – Kohonen-Karten.
<https://www.geeksforgeeks.org/self-organising-maps-kohonen-maps/>, 2023

Abbildung 87: Deep Belief Networks
Grellmann, Martin. 10 Deep Learning Algorithmen, die Sie kennen sollten.
<https://martin-grellmann.de/10-deep-learning-algorithmen-die-sie-kennen-sollten>, 2021

Abbildung 88: Restricted Boltzmann Machines
Grellmann, Martin. 10 Deep Learning Algorithmen, die Sie kennen sollten.
<https://martin-grellmann.de/10-deep-learning-algorithmen-die-sie-kennen-sollten>, 2021

Oppermann, Artem. Deep Learning Trifft auf Physik: Restricted Boltzmann Machines.
<https://artemoppermann.com/de/deep-learning-trifft-auf-physik-restricted-boltzmann-machines/>

Abbildung 89: Autoencoders
Grellmann, Martin. 10 Deep Learning Algorithmen, die Sie kennen sollten.
<https://martin-grellmann.de/10-deep-learning-algorithmen-die-sie-kennen-sollten>, 2021

Abbildung 90: Machine Learning nutzt Daten, um Muster und Zusammenhänge in Daten zu identifizieren
datasolut, Wuttke, Vincent. Machine Learning: Definition, Algorithmen, Methoden und Beispiele.
<https://datasolut.com/was-ist-machine-learning/#unueberwachtes-lernen>, 2024

Abbildung 91: Lineare Regression-Algorithmus
AWS. Was ist der Unterschied zwischen linearer Regression und logistischer Regression?
<https://aws.amazon.com/de/compare/the-difference-between-linear-regression-and-logistic-regression/>

Abbildung 92: *Logistische Regression-Algorithmus*

Microsoft. Machine-Learning-Algorithmen - Eine Einführung in die Mathematik und Logik von Machine Learning.

<https://azure.microsoft.com/de-de/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms>

Abbildung 93: *Vergleich lineare Regression vs. logistische Regression*

awa. Was ist der Unterschied zwischen linearer Regression und logistischer Regression?

<https://aws.amazon.com/de/compare/the-difference-between-linear-regression-and-logistic-regression/>

Abbildung 94: *Naïve Bayes-Naive Bayes-Klassifikatoren-Algorithmus*

Microsoft. Machine-Learning-Algorithmen - Eine Einführung in die Mathematik und Logik von Machine Learning.

<https://azure.microsoft.com/de-de/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms>

Abbildung 95: *Support Vector Machine-Algorithmus (SVM) Algorithmus*

Beneker, Daniel. Algorithmus im Detail.

<https://fh-bielefeld-mif-sw-engineerin.gitbooks.io/script/content/ai/support-vector-machine/algorithmus-im-detail.html>

Abbildung 96: *Entscheidungsstruktur-Algorithmen*

Microsoft. Machine-Learning-Algorithmen - Eine Einführung in die Mathematik und Logik von Machine Learning.

<https://azure.microsoft.com/de-de/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms>

Abbildung 97: *KNN-Diagramm*

IBM. Was ist der K-Nächste-Nachbarn- Algorithmus?

<https://www.ibm.com/de-de/topics/knn>

Abbildung 98: *Clustering-Algorithmus*

Vedder, Marcel. Clusteranalyse einfach erklärt.

<https://blog.enra.app/clusteranalyse-einfach-erklart/>, 2021

Abbildung 99: *k-Means Clustering Prozess*

DATA BASE CAMP. Was ist k-Means Clustering?

<https://databasecamp.de/ki/k-means-cluster>, 2022

Abbildung 100: *Ergebnisse unseres Clusteranalyse-Beispiels. Clusterbildung mit dem DBSCAN-Algorithmus. Auswertung der gefundenen Cluster mit dem Calinski-Harabasz-Index und der Silhouttenmethode*

Marzell, Tobias. Clustering mit Machine Learning - Ein ausführlicher Leitfaden.

<https://rocketloop.de/de/blog/clustering-machine-learning-ausfuhrlicher-leitfaden/>, 2021

Abbildung 101: *Clusterbildung mit dem HDBSCAN-Algorithmus. Auswertung der gefundenen Cluster mit dem Calinski-Harabasz-Index und der Silhouttenmethode*

Marzell, Tobias. Clustering mit Machine Learning - Ein ausführlicher Leitfaden.

<https://rocketloop.de/de/blog/clustering-machine-learning-ausfuhrlicher-leitfaden/>, 2021

Abbildung 102: *beispielhafte Darstellung eines hierarchischen Clusterings beim Machine Learning*

Marzell, Tobias. Clustering mit Machine Learning - Ein ausführlicher Leitfaden.

<https://rocketloop.de/de/blog/clustering-machine-learning-ausfuhrlicher-leitfaden/>, 2021

Abbildung 103: Random-Forest-Algorithmus

Hemashreekilari. Understanding Random Forest.

<https://medium.com/@hemashreekilari9/understanding-random-forest-a87d08416280>, 2023

Abbildung 104: AdaBoost

AlmaBetter. AdaBoost Algorithm in Machine Learning.

<https://www.almabetter.com/bytes/tutorials/data-science/adaboost-algorithm>, 2023

Abbildung 105: Gradient Boosting-Algorithmus

Hemashreekilari. Understanding Gradient Boosting.

<https://medium.com/@ilyurek/light-gbm-a-powerful-gradient-boosting-algorithm-fe145a1cd8a6>, 2023

Abbildung 106: LightGBM

Kilic, Ilyurek. Light GBM: A Powerful Gradient Boosting Algorithm.

<https://medium.com/@ilyurek/light-gbm-a-powerful-gradient-boosting-algorithm-fe145a1cd8a6>, 2023

Abbildung 107: CatBoost

Shahani, Muhammad Niaz; Zheng, Xigui; Guo, Xiaowei; Wie, Xin. Machine Learning-Based Intelligent Prediction of Elastic Modulus of Rocks at Thar Coalfield.

https://www.researchgate.net/figure/Explanation-of-the-Catboost_fig6_359380024, 2023

Abbildung 108: XGBoost

Zhao, Zhiqian; Wang, Tao; Gao, Dianrong. Degradation state recognition of piston pump based on ICEEMDAN and XGBoost.

https://www.researchgate.net/figure/Flow-chart-of-XGBoost_fig3_345327934, 2020

Abbildung 109: Deep neural network

IBM. AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference?

<https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>, 2023

Abbildung 110: Machine Learning vs. Deep Learning: der Unterschied liegt in der Feature Extraktion und dem Einsatz von tiefen, künstlichen neuronalen Netzen

datasolut. Wuttke, Vinzent. Machine Learning: Definition, Algorithmen, Methoden und Beispiele.

<https://datasolut.com/was-ist-machine-learning/#unueberwachtes-lernen>, 2024

Abbildung 111: Natural Language Processing

Wuttke, Laurenz. NLP vs. NLU vs. NLG: Unterschiede, Funktionen und Beispiele.

<https://datasolut.com/natural-language-processing-vs-nlu-vs-nlg-unterschiede-funktionen-und-beispiele/>, 2023

Abbildung 112: Unterschiede von NLP, NLU und NLG

Wuttke, Laurenz. NLP vs. NLU vs. NLG: Unterschiede, Funktionen und Beispiele.

<https://datasolut.com/natural-language-processing-vs-nlu-vs-nlg-unterschiede-funktionen-und-beispiele/>, 2023

Abbildung 113: Datenerhebung von Bais

Klier, Prof. Dr. Mathias. Grundlagen. Zu Bias & Fairness in KI-Systemen.

<https://bias-and-fairness-in-ai-systems.de/grundlagen/>, 2024

Abbildung 114: Entwicklung, Implementierung und Nutzung von Bais
Klier, Prof. Dr. Mathias. Grundlagen. Zu Bias & Fairness in KI-Systemen.
<https://bias-and-fairness-in-ai-systems.de/grundlagen/>, 2024

Abbildung 115: Formaler Aufbau einer KI-Anwendung
Poretschkin, Dr. Maximilian; Schmitz, Anna; Akila, Dr. Maram; Adilova, Linara; Becker, Dr. Daniel; Cremers, Prof. Dr. Armin B.; Hecker, Dr. Dirk; Houben, Dr. Sebastian; Mock, PD Dr. Michael; Rosenzweig, Julia; Sicking, Joachim; Schulz, Elena; Voss, Dr. Ang. Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz - KI-Prüfkatalog.
https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatalog/202107_KI-Pruefkatalog.pdf, 2021, S.18

Abbildung 116: Lebenszyklus einer KI-Anwendung
Klier, Prof. Dr. Mathias. Grundlagen. Zu Bias & Fairness in KI-Systemen.
<https://bias-and-fairness-in-ai-systems.de/grundlagen/>, 2024

Abbildung 117: Abstrahierter Lebenszyklus einer KI-Anwendung
Klier, Prof. Dr. Mathias. Grundlagen. Zu Bias & Fairness in KI-Systemen.
<https://bias-and-fairness-in-ai-systems.de/grundlagen/>, 2024

Abbildung 118: Training des ML-Modells einer KI-Anwendung
Poretschkin, Dr. Maximilian; Schmitz, Anna; Akila, Dr. Maram; Adilova, Linara; Becker, Dr. Daniel; Cremers, Prof. Dr. Armin B.; Hecker, Dr. Dirk; Houben, Dr. Sebastian; Mock, PD Dr. Michael; Rosenzweig, Julia; Sicking, Joachim; Schulz, Elena; Voss, Dr. Ang. Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz - KI-Prüfkatalog.
https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatalog/202107_KI-Pruefkatalog.pdf, 2021, S.21

Abbildung 119: IOT & Maschinelles Lernen in 3 Stufen ML-Ansatz
Nieweg, Hendrik. Whithpaper. Künstliche Intelegenz in der IoT-Praxis - Use Case und Erfolgsfaktoren.
https://device-insight.com/wp-content/uploads/2022/04/180919-DeviceInsight-Whitepaper-DE-WEB.pdf?utm_campaign=Follow-up%20Mails&utm_medium=email&_hsenc=p2ANqtz--BJgfFxn9jr18D9bn15l2P-P7w4ocNnKVDcf1C1f6RSIjTpZIO-SshmkHABiyX46bpmOZoQpJsgn6XVFIEUuavzYQTg&_hsmi=88840290&utm_content=88840290&utm_source=hs_automation&hsCtaTracking=ef7f4df9-81d1-40c3-8135-1988131c996d%7C40b782a5-ff39-44aa-abad-ea777d79442c, S.9

Abbildung 120: wesentliche Unterscheidungsmerkmale zwischen SIEM und SOAR
Houdeau, Detlef. Wie kann KI vor Cyberangriffen schützen?
<https://crisis-prevention.de/kommunikation-it/wie-kann-ki-vor-cyberangriffen-schuetzen.html#>, 2023

Abbildung 121: Rapid Connectivity Installation and Asset Onboarding
Nieweg, Hendrik. Whithpaper. Künstliche Intelegenz in der IoT-Praxis- Use Case und Erfolgsfaktoren.
https://device-insight.com/wp-content/uploads/2022/04/180919-DeviceInsight-Whitepaper-DE-WEB.pdf?utm_campaign=Follow-up%20Mails&utm_medium=email&_hsenc=p2ANqtz--BJgfFxn9jr18D9bn15l2P-P7w4ocNnKVDcf1C1f6RSIjTpZIO-SshmkHABiyX46bpmOZoQpJsgn6XVFIEUuavzYQTg&_hsmi=88840290&utm_content=88840290&utm_source=hs_automation&hsCtaTracking=ef7f4df9-81d1-40c3-8135-1988131c996d%7C40b782a5-ff39-44aa-abad-ea777d79442c, S.8

Abbildung 122: Responsible AI Principles von infotech

INFO-TECH Research Group. Develop Responsible AI Guiding Principles

Find your north star for responsible AI.

<https://www.infotech.com/research/ss/develop-responsible-ai-guiding-principles>

Abbildung 123: Laufe des AI-Acts von 2024 bis 2026

Retting, Julia; Müller, Fabian. AI. ACT Quick Check - We creat the next.

https://www.statworx.com/ai-act/tool/Ergebnis_AI-Act-Quick-Check_nicht-betroffen.pdf, S.3

Abbildung 124: KI-Regulierungen und ihre Umsetzungen

PWC. Reese, Hendrik. EU AI Act: Europäische KI-Regulierung und ihre Umsetzung. [https://](https://www.pwc.de/de/risk-regulatory/responsible-ai/europaeische-ki-regulierung-und-ihre-umsetzung.html)

www.pwc.de/de/risk-regulatory/responsible-ai/europaeische-ki-regulierung-und-ihre-umsetzung.html

Abbildung 125: Hochrisiko-KI-Systeme nach Annex I und III

PWC. Reese, Hendrik. EU AI Act: Europäische KI-Regulierung und ihre Umsetzung. [https://](https://www.pwc.de/de/content/16cb46ed-314e-44b5-881e-6a252d79ae46/pwc-whitepaper-eu-ai-act.pdf)

www.pwc.de/de/content/16cb46ed-314e-44b5-881e-6a252d79ae46/pwc-whitepaper-eu-ai-act.pdf, 2024. S.8

Abbildung 126: KI-Risiko Pyramiede

Dreyer, Schürmann Rosenthal. KI-Verordnung kommt: Leitfaden für Unternehmen. [https://](https://www.srd-rechtsanwaelte.de/ki-verordnung/#block-request-form)

www.srd-rechtsanwaelte.de/ki-verordnung/#block-request-form, SRD_Whitepaper_KI-VO.pdf (gesendetes .pdf) , S.17

Abbildung 127: Kategorien von KI-Systemen nach dem EU AI-Act

PWC. Vertrauenswürdige KI. Umsetzung des EU AI Act als Value Treiber.

<https://www.pwc.de/de/content/16cb46ed-314e-44b5-881e-6a252d79ae46/pwc-whitepaper-eu-ai-act.pdf>, 2024. S.7

Abbildung 128: gesetzliche KI-Verordnung der EU

Dreyer, Schürmann Rosenthal. KI-Verordnung kommt: Leitfaden für Unternehmen. [https://](https://www.srd-rechtsanwaelte.de/ki-verordnung/#block-request-form)

www.srd-rechtsanwaelte.de/ki-verordnung/#block-request-form, SRD_Whitepaper_KI-VO.pdf (gesendetes .pdf) , S.26

Abbildung 129: internationale Landschaft der KI-Initiativen

Europäischen Kommission. Studie zur Unterstützung der Folgenabschätzung zur KI-Verordnung.

<https://digital-strategy.ec.europa.eu/de/library/study-supporting-impact-assessment-ai-regulation>, Study supporting the impact assessment of the AI regulation.pdf, 2023

Abbildung 130: Risk Mitigation Essential

INFO-TECH. Rohde, Logan. Address Security and Privacy Risk for Generative AI.

<https://www.infotech.com/research/ss/address-security-and-privacy-risks-for-generative-ai>

Abbildung 131: Lebenszyklus des Analysemodells

Baquero, Juan Aristi; Burkhardt, Roger; Govindaraja, Arvind; Wallace, Thomas.

Derisking AI by design: How to bulid risk managemant into AI development.

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development>, 2020

Abbildung 132: Risikoentwicklung

Baquero, Juan Aristi; Burkhardt, Roger; Govindaraja, Arvind; Wallace, Thomas.

Derisking AI by design: How to build risk management into AI development.

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development>, 2020

Abbildung 133: Risikoklassifizierung

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 134: Risk Management Framework

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 135: Risk Management Framework

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 136: 2 Level Algorithmic Risk Monitor und Management

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 137: ML/AI basiert extrinsisches Risikomanagement

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 138: Algorithmische Risikobewertung und Auswirkungsabschätzung

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 139: Risk Governance

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 140: Künstliche Entscheidungsfreiheit als Grundlage für risikobasierte Governance

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>, 2022

Abbildung 141: -KI-Analyse

aus selbsterstellten Excel File: **SOC-AI.xlsx**

Abbildung 142: AI-Act-Analyse

aus selbsterstellten Excel File: **AI-Act-EU.xlsx**

Abbildung 143: KI-Eigenschaften, die auf Richtliniendokumente abgebildet werden
Nineta Polemi, Isabel Praça. enisa. A multilayer framework for good cybersecurity practices for AI.
<https://www.cybersecitalia.it/wp-content/uploads/2023/06/Multilayer-Framework-for-Good-Cybersecurity-Practices-for-AI.pdf>, 2023, S.21

Abbildung 144: Beziehung zwischen KI-Bedrohungen und Sicherheitskontrollen
Nineta Polemi, Isabel Praça. enisa. A multilayer framework for good cybersecurity practices for AI.
<https://www.cybersecitalia.it/wp-content/uploads/2023/06/Multilayer-Framework-for-Good-Cybersecurity-Practices-for-AI.pdf>, 2023, S.22

Abbildung 145: Vergleich klassische Computer vs. Quantenrechner
Klasen, Florenz. IT-Trends 2024: Die 15 wichtigsten Technologien der Zukunft.
<https://techminds.de/magazin/it-trends/>, 2023

Abbildung 146: Übersicht von der Cyber-Security
Vieth, Simon. Cyber Security – Welche Fragen der IT-Sicherheit bewegen den Markt?
<https://www.conet.de/blog/cyber-security-welche-fragen-der-it-sicherheit-bewegen-aktuell-den-markt/>, 2022

Abbildung 147: NKCS-Verbund
BSI. Nationales Koordinierungszentrum für Cybersicherheit (NKCS).
<https://www.forschung-it-sicherheit-kommunikationssysteme.de/forschung/it-sicherheit/nkcs>

Abbildung 148: Übersicht über BSI-Publikationen zum Sicherheitsmanagement
BSI. BSI-Standart 200-1. Managementsysteme für Informationssicherheit (ISMS).
https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/BSI_Standards/standart_200_1.pdf?__blob=publicationFile&v=2, S.11

Abbildung 149: Aufbau der Energiesynchronisationsplattform zur Automatisierung und Standardisierung des Energieflexibilitätshandels
Synergie. INFORMATIONS- UND KOMMUNIKATIONSTECHNIK.
<https://synergie-projekt.de/ueber-synergie/arbeitsgebiete/synergie-1-und-2/informations-und-k>

Abbildung 150: Key Concepts of COBIT 5
Nash V.. Key Concepts and Principles of COBIT 5 Explained.
<https://www.itsm-docs.com/en-nl/blogs/cobit/cobit-5-key-concepts>, 2023

Abbildung 151: ITIL-Zertifizierung von Service-Management-Systemen
ITIL®. New ITIL® 4 Qualification Scheme – ITIL® 4 Master.
<https://agilizing.com/new-til-4-qualification-scheme-til-4-m>, 2023

Abbildung 152: NIST Cyber Security Framework 2.0
INFOSECTRAIN. NIST Cybersecurity Framework 2.0.
<https://www.infosectrain.com/blog/nist-cybersecurity-framework/>, 2023

Abbildung 153: DISA-Überblick
FB PRO GMBH: CYBERSECURITY, SYSTEMHÄRTUNG UND MEHR: DIESE AUFGABEN HABEN BSI, DISA, ACSC UND CIS.
<https://www.fb-pro.com/cybersecurity-organisationen-behoerden/>

Abbildung 154: CIS

FB PRO GMBH: CYBERSECURITY, SYSTEMHÄRTUNG UND MEHR: DIESE AUFGABEN HABEN BSI, DISA, ACSC UND CIS.

<https://www.fb-pro.com/cybersecurity-organisationen-behoerden/>

Abbildung 155: ACSC-Logo

ATLASSIAN. ACSC: Essential Eight Maturity Model (Maturitätsmodell "Essential Eight").

<https://www.atlassian.com/de/trust/compliance/resources/essential8>, 2024

Abbildung 156: BSI-Schutzziele ISO/IEC 27001

Peterjohann, Horst. Informationssicherheit. - Die Schutzziele Vertraulichkeit, Integrität und Verfügbarkeit sicherstellen.

<https://www.peterjohann-consulting.de/informationssicherheit/>, 2023

Abbildung 157: Gefährdungen - Schutzziele – Schutzbedarfe

DriveLock. Vertraulichkeit, Integrität und Verfügbarkeit: von IT Schutzziele zu konkreten Maßnahmen.

<https://www.drivelock.com/de/blog/vertraulichkeit-integritaet-verfuegbarkeit-schutzziele-bsi-grundschutz>, 2021

Abbildung 158: Risikomatrix mit Risikoeinstufung

Dieter Maul-Burton. IT Security Risikoanalyse vermeidet Kosten.

<https://adiccon.de/it-security-risikoanalyse-vermeidet-kosten/>, 2016

BSI. Lerneinheit 7.7: Risiken bewerten.

https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/Zertifizierte-Informationssicherheit/IT-Grundschutzschulung/Online-Kurs-IT-Grundschutz/Lektion_7_Risikoanalyse/Lektion_7_07/Lektion_7_07_node.html

(Grafik wurde zusammengestellt)

Abbildung 159: Zusammenhang zwischen Angriffen auf die Schutzziele und Gegenmaßnahmen

Stemplewitz, Thomas. Konzeption von IT-Sicherheitskriterien für vernetzte Endgeräte.

https://it-forensik.fiw.hs-wismar.de/images/a/aa/MT_Stemplewitz.pdf, 2019, S.19

Abbildung 160: Elemente einer Cyber-Security-Strategie

Lünendonk & Hossenfelder GmbH. Cyber Security ist das Top-Thema für CIOs.

<https://www.luenendonk.de/aktuelles/presseinformationen/cyber-security-ist-das-top-thema-fuer-cios/>, 2021

Abbildung 161: Top 10 Cyberangriffe 2023

Prajwal. List Of Top Cybersecurity Threats In 2024.

<https://www.sprintzeal.com/blog/top-cybersecurity-threats>, 2023

Abbildung 162: Die Lage der IT-Sicherheit in Deutschland 2023 im Überblick und Bedrohungsziele laut BSI

Kröger, Janina. BSI-Lagebericht 2023. Bedrohung im Cyberraum so hoch wie nie zuvor.

<https://it-service.network/blog/2023/11/03/bsi-lagebericht-2023/>, 2023

Abbildung 163: Threat – Asset – Vulnerability – Risk - Zusammenhänge

COGNUSSYSTEMS. Our Vulnerability & Risk assessments.

<https://cognussys.com/vulnerability-and-risk-assessments/>

Anhang VI Tabellenverzeichnis

Excel File: NIST-SOC-AI.xlsx

Excel File: Act-EU.xlsx

Anhang VII Literaturverzeichnis

Acronis. *Unverzichtbare Komponenten bei modernem und zuverlässigem Ransomware-Schutz.*

<https://www.acronis.com/de-de/blog/posts/essential-components-for-robust-ransomware-protection-in-the-modern-age/>, 2023

AI Universe. *KI-Tools.*

<https://www.ai-universe.com/ki-tools/>, 2024

AIGA. *List of AI Governance Tasks.*

<https://ai-governance.eu/ai-governance-framework/task-list/> ff., 2023

Ashwani, K.. *What is LogRhythm and use cases of LogRhythm?*

<https://www.devopsschool.com/blog/what-is-logrhythm-and-use-cases-of-logrhythm/>, 2023

AWS. *Was ist der Unterschied zwischen linearer Regression und logistischer Regression?*

<https://aws.amazon.com/de/compare/the-difference-between-linear-regression-and-logistic-regression/>

BIAS & FAIRNESS. Institut für Business Analytics. *Grundlagen. Zu Bias & Fairness in KI-Systemen.*

<https://bias-and-fairness-in-ai-systems.de/grundlagen/>, 2024

Baquero, Juan Aristi; Burkhardt, Roger; Govindaraja, Arvind; Wallace, Thomas.

Derisking AI by design: How to build risk management into AI development.

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development>, 2020

BSI.

https://www.bsi.bund.de/DE/Home/home_node.html

BSI. *IT-Grundschatz-Bausteine.*

https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschatz/IT-Grundschatz-Kompendium/IT-Grundschatz-Bausteine/Bausteine_Download_Edition_node.html

BSI. *Nationales Koordinierungszentrum für Cybersicherheit (NKCS).*

<https://www.forschung-it-sicherheit-kommunikationssysteme.de/forschung/it-sicherheit/nkcs>

BSI. *BSI-Magazin 2023/02. Mit Sicherheit.*

https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Magazin/BSI-Magazin_2023_02.pdf?__blob=publicationFile&v=4,%20Februar%202023, 2023

BSI. *BSI-Standard 200-2. Managementsysteme für Informationssicherheit (ISMS).*

https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschatz/BSI_Standards/standard_200_1.pdf?__blob=publicationFile&v=2, 2023

Clickworker. *Generative Adversarial Networks (GANs).*
<https://www.clickworker.de/ki-glossar/generative-adversarial-networks/>

CLOUDFLARE. *Was ist ein neuronales Netzwerk?*
<https://www.cloudflare.com/de-de/learning/ai/what-is-neural-network/>

COO, Frederic Noppe. *Ransomware: Ablauf eines Angriffs und Gegenmassnahmen.*
<https://l3montree.com/publikationen/ransomware-ablauf-eines-angriffs-und-gegenmassnahmen>

csialtd. *business-continuity-and-scenario-template.doc.*
https://csialtd.com.au/wp-content/uploads/2020/05/business-continuity-and-scenario-template.docx?_cf_chl_tk=OOOqzKo_vGc8Dr9l3qxdkvVI8PVOI8lay2UqttvOu10-1707338540-0-zQr7

DATA BASE CAMP. *Was ist k-Means Clustering?*
<https://databasecamp.de/ki/k-means-cluster>, 2022

Dreyer, Schürmann Rosenthal. *KI-Verordnung kommt: Leitfaden für Unternehmen.*
<https://www.srd-rechtsanwaelte.de/ki-verordnung/#block-request-form>, SRD_Whitepaper_KI-VO.pdf (gesendetes .pdf)

Exabeam. *What Is Log Aggregation? The Complete Guide.*
<https://www.exabeam.com/explainers/event-logging/log-aggregation/>

European Union. 2024/1689 . VERORDNUNG (EU) 2024/1689 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz)
https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401689, 2024

European Union. *NIS2-Richtlinie.*
<https://digital-strategy.ec.europa.eu/de/node/10361/printable/pdf>, 2024

European Union Agency for Cybersecurity. Polemi, Nineta; Praça, Isabel. *A multilayer framework for good cybersecurity practices for AI.*
<https://www.cybersecitalia.it/wp-content/uploads/2023/06/Multilayer-Framework-for-Good-Cybersecurity-Practices-for-AI.pdf>, 2023

European Union. *VERORDNUNG DES EUROPÄISCHEN PARLAMENTS UND DES RATES ZUR FESTLEGUNG HARMONISierter VORSCHRIFTEN FÜR KÜNSTLICHE INTELLIGENZ (GESETZ ÜBER KÜNSTLICHE INTELLIGENZ) UND ZUR ÄNDERUNG BESTIMMTER RECHTSAKTE DER UNION.*
https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_1&format=PDF, 2021

European Union. *ZUR FESTLEGUNG HARMONISierter VORSCHRIFTEN FÜR KÜNSTLICHE INTELLIGENZ (GESETZ ÜBER KÜNSTLICHE INTELLIGENZ) UND ZUR ÄNDERUNG BESTIMMTER RECHTSAKTE DER UNION.*
https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_2&format=PDF, 2021

European Union. ANHÄNGE des Vorschlags für eine Verordnung des Europäischen Parlaments und des Rates.

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_2&format=PDF, 2021

European Union. Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe FINAL REPORT (D5). Study supporting the impact assessment of the AI regulation.pdf

<https://digital-strategy.ec.europa.eu/de/library/study-supporting-impact-assessment-ai-regulation>, 2023

European Union. NIS2. RICHTLINIE (EU) 2022/2555 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 14. Dezember 2022.

<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32022L2555>, 2022

FB Pro GmbH. System Hardening & Data Protection. Cybersecurity, Systemhärtung und mehr: Diese Aufgaben haben BSI, DISA, ACSC und CIS.

<https://www.fb-pro.com/cybersecurity-organisationen-behoerden/>

Fraunhofer IAIS. Fraunhofer-Institut für Intelligente Analyse-und Informationssysteme IAIS. Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz.

<https://www.iais.fraunhofer.de/s/ki-pruefkatalog/index.html>, 2021

Grellmann, Martin. 10 Deep Learning Algorithmen, die Sie kennen sollten.

<https://martin-grellmann.de/10-deep-learning-algorithmen-die-sie-kennen-sollten>, 2021

Geeksforgeeks. Gradient Boosting in ML.

<https://www.geeksforgeeks.org/ml-gradient-boosting/>, 2023

GoThesis. Plagiatsprüfung inkl. KI-Erkennung

<https://www.gothesis.de/>

Gupta, Somil. Algorithmic Risk Management: A Framework for Identifying, Assessing, Controlling, and Mitigating Risks in AI Development and Operations.

<https://hyperight.com/algorithmic-risk-management/>

Hecker Consulting. Die 10 wichtigsten IT-Trends in 2024 die man kennen sollte.

<https://www.hco.de/blog/die-10-wichtigsten-it-trends-in-2024-die-man-kennen-sollte>, 2023

Houdea, Detlef. Wie kann KI vor Cyberangriffen schützen?

<https://crisis-prevention.de/kommunikation-it/wie-kann-ki-vor-cyberangriffen-schuetzen.html#>, 2023

Huber, Hermann. Wie kann KI bei der Bekämpfung von Phishing-Attacken helfen?

<https://de.linkedin.com/pulse/wie-kann-ki-bei-der-bek%C3%A4mpfung-von-phishing-attacken-helfen-huber-yok0e>, 2023

IBM. IBM Security QRadar Suite.

<https://www.ibm.com/de-de/products/qradar-siem>

IBM. IBM Security QRadar Suite.

<https://www.ibm.com/de-de/qradar>

IBM. Network Detection and Response (NDR).

<https://www.ibm.com/de-de/products/qradar-siem/ndr>

IBM. QRadar content extensions.

<https://www.ibm.com/docs/en/qsip/7.4?topic=qradar-content-extensions>, 2024

IBM. QRadar architecture overview.

<https://chat.openai.com/c/462c7c70-e58c-4fd4-898c-32279bb71214>, 2023

IBM. Metric-based machine learning overview.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-metric-based-machine-learning>, 2024

IBM. Z Topology integration.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-z-topology-integration>, 2024

IBM. Problem insights overview.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-problem-insights>, 2024

IBM. Metric-based machine learning overview.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-metric-based-machine-learning>, 2024

IBM. Log-based machine learning overview.

<https://www.ibm.com/docs/en/z-anomaly-analytics/5.1.0?topic=overview-log-based-machine-learning>, 2024

IBM. The AI and data platform that's built for business.

<https://www.ibm.com/watsonx?lnk=flatitem>

IBM. IBM Security QRadar SIEM demo.

https://mediacenter.ibm.com/media/IBM+Security+QRadar+SIEM+demo/1_ygg5ijnj

IBM. Was sind neuronale Netze?

<https://www.ibm.com/de-de/topics/neural-networks>

IBM. IBM Data and AI Team. AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference?

<https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>, 2023

IBM. Was ist der K-Nächste-Nachbarn- Algorithmus?

<https://www.ibm.com/de-de/topics/knn>

IBM. Was ist Random-Forest-Algorithmus?

<https://www.ibm.com/de-de/topics/random-forest>

IBM. Entscheidungsstrukturen.

<https://www.ibm.com/docs/de/cognos-analytics/11.1.0?topic=pada-decision-tree>, 2024

IBM. Was sind naive Bayes-Klassifikatoren?

<https://www.ibm.com/de-de/topics/naive-bayes>

Instituts für Business Analytics (IBA). Grundlagen. Zu Bias & Fairness in KI-Systemen.

<https://bias-and-fairness-in-ai-systems.de/grundlagen/>, 2023

IQWIG. Biasarten.

<https://www.iqwig.de/sonstiges/glossar/biasarten.html>

Jupyter.

<https://jupyter.org>

Kanade, Vijay. *What Is a Support Vector Machine? Working, Types, and Examples.*

<https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/>, 2022

Kröger, Janina. *BSI-Lagebericht 2023. Bedrohung im Cyberraum so hoch wie nie zuvor.*

<https://it-service.network/blog/2023/11/03/bsi-lagebericht-2023/>, 2023

Kruger, Nicolene. *Managing AI Governance in the Future: An Overview of the EU AI Act, ISO/IEC 42001, and NIST AI RMF.*

<https://certpro.com/ai-governance-eu-iso-nist-overview>, 2024

Keary, Tim. *The Best SIEM Tools for 2024: Vendors & Solutions Ranked.*

<https://www.comparitech.com/net-admin/siem-tools/>, 2024

Klasen, Florenz. *IT-Trends 2024: Die 15 wichtigsten Technologien der Zukunft.*

<https://techminds.de/magazin/it-trends/>, 2023

KOGIT. *Security Information & Event Management.*

<https://www.kogit.de/services/security-information-event-management/>

KOGIT. *ECHTZEITANALYSEN VON SICHERHEITSRELEVANTEN VORFÄLLEN MIT SIEM.*

<https://www.kogit.de/services/security-information-event-management/>

LaPiedra, James. *Global Information Assurance Certification Paper.*

<https://www.giac.org/paper/gsec/501/information-security-process-prevention-detection-response/101197>, 2002

LightGBM. *Welcome to LightGBM's documentation!*

<https://lightgbm.readthedocs.io/en/latest/>

Logpoint. *SIEM reduziert das Cyber-Risiko mit leistungsstarken Datenanalysen.*

<https://www.logpoint.com/de/produkt/logpoint-als-siem-werkzeug/>

Logpoint. *Die Top 10 SIEM-Use-Cases für Ihre Implementierung.*

<https://www.logpoint.com/de/die-10-wichtigsten-siem-anwendungsfalle/>

Logpoint. *SIEM Buyer's Guide 2023.*

<https://www.logpoint.com/wp-content/uploads/2023/05/siem-buyers-guide-2023.pdf>, 2023

Logpoint. *Was ist User and Entity Behavior Analytics? Ein umfassender Leitfaden zu UEBA, der Funktionsweise und den Vorteilen.*

<https://www.logpoint.com/de/blog/what-is-ueba-a-complete-guide-to-ueba/>, 2020

Logpoint. *LogPoint Use Cases.*

<https://www.logpoint.com/wp-content/uploads/2022/03/logpoint-use-cases-whitepaper.pdf>

Logpoint. Bogati, Anish. *DarkGate infection chain.*

<https://www.logpoint.com/en/blog/inside-darkgate/>, 2024

Logpoint. *SIEM reduziert das Cyber-Risiko mit leistungsstarken Datenanalysen.*

<https://www.logpoint.com/de/produkt/logpoint-als-siem-werkzeug/>

LogRhythm.

<https://logrhythm.com/>

LogRhythm. *Maschinelles Lernen in der IT-Sicherheit: Ein datenwissenschaftlich orientierter Ansatz.*

<https://gallery.logrhythm.com/white-papers-and-e-books/de-employing-machine-learning-in-a-security-environment-white-paper.pdf>

LogRhythm. *Entdecken Sie das LogRhythm NextGen SIEM.*

[https://logrhythm.com/de-schedule-online-demo/?utm_medium=cpc&utm_source=Google&utm_campaign=LogRhythm_CEUR_Brand_Pure_T1&utm_term={Adgroup}=EMEA&utm_region=EMEA?&utm_medium=cpc&utm_source=Google&utm_campaign=LogRhythm-DE\(N\)-T1-BrandPure&utm_term=Brand](https://logrhythm.com/de-schedule-online-demo/?utm_medium=cpc&utm_source=Google&utm_campaign=LogRhythm_CEUR_Brand_Pure_T1&utm_term={Adgroup}=EMEA&utm_region=EMEA?&utm_medium=cpc&utm_source=Google&utm_campaign=LogRhythm-DE(N)-T1-BrandPure&utm_term=Brand)

Management Circle. *DAS GROSSE 1X1 DER KÜNSTLICHEN INTELLIGENZ.*

<https://go.managementcircle.de/ki-1x1>

ManageEngine. *Different types of logs in SIEM and their log formats.*

<https://www.manageengine.com/log-management/siem/collecting-and-analysing-different-log-types.html>

ManageEngine. *Take control of your IT.*

<https://www.manageengine.com/>

ManageEngine. *Komplettes Endpoint Management – über ein einziges Interface.*

https://www.manageengine.de/fileadmin/user_upload/02_Produkte-Loesungen/Desktop_Central/Endpoint-Central-Datenblatt.pdf, 2024

ManageEngine. *A single pane of glass for complete Endpoint Management and Security.*

<https://download.manageengine.com/products/desktop-central/desktop-administration-overview.pdf>, 2024

ManageEngine. *Enterprise AIOps. How ManageEngine refines IT processes with artificial intelligence.*

https://download.manageengine.com/academy/aops_e-book.pdf

ManageEngine. *OpManager Online Demo*

<https://demo.opmanager.com/>

Mäntymäki, Matti; Minkkinen, Matti; Birkstedt, Teemu, and Viljanen, Mika.

Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance.

<https://arxiv.org/pdf/2206.00335>, 2023

Marzell, Tobias. *Clustering mit Machine Learning - Ein ausführlicher Leitfaden.*

<https://rocketloop.de/de/blog/clustering-machine-learning-ausfuhrlicher-leitfaden/>, 2021

McCartney, Ava. *Die 10 wichtigsten strategischen Technologie-Trends von Gartner für 2024.*

<https://www.gartner.de/de/artikel/die-10-wichtigsten-strategischen-technologie-trends-von-gartner-fuer-2024>, 2023

Miloslavskaya, Natalia. *Analysis of SIEM Systems and Their Usage in Security Operations and Security Intelligence Centers.*

https://www.researchgate.net/publication/318708872_Analysis_of_SIEM_Systems_and_Their_Usage_in_Security_Operations_and_Security_Intelligence_Centers, 2018

Miuccio, Tony. *SIEM Maturity Hierarchy.*
<https://blacktowersec.com/siem-maturity-hierarchy/>

Mohan, Remya. *What Is Security Information and Event Management (SIEM)? Definition, Architecture, Operational Process, and Best Practices. SIEM manages vulnerabilities by providing next-generation detection, analytics, and Best Practices. SIEM manages vulnerabilities by providing next-generation detection, analytics, and response.*
<https://www.spiceworks.com/it-security/vulnerability-management/articles/what-is-siem/>

NIEWEG, HENDRIK. *Künstliche Intelligenz in der IoT-Praxis – Use Cases und Erfolgsfaktoren.*
https://device-insight.com/wp-content/uploads/2022/04/180919-DeviceInsight-Whitepaper-DE-WEB.pdf?utm_campaign=Follow-up%20Mails&utm_medium=email&_hsenc=p2ANqtz--BJgfFxn9jr18D9bn15I2P-P7w4ocNnKVDcf1C1f6RSIjTpZIO-SshmkHABiyX46bpmOZoQpJsgn6XVFiEUuavzYQTg&_h

NIS2.
<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32022L2555>

NIST.
<https://www.nist.gov/>

NIST. *NIST AI 100-1 Artificial Intelligence Risk Management Framework (AI RMF 1.0).*
<https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>, 2023,

NIST. *National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0).*
<https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>

RCDev. *Erkundung von 7 neuen KI-gestützten Cyber-Bedrohungen und wie die Software von RCDevs wirksamen Schutz bieten kann.*
<https://www.rcdevs.com/de/exploring-7-new-ai-powered-cyber-threats-and-how-rcdevs-software-can-provide-effective-protection>, 2023

Reese, Hendrik. *EU AI Act: Europäische KI-Regulierung und ihre Umsetzung.*
https://www.pwc.de/de/risk-regulatory/responsible-ai/europaeische-ki-regulierung-und-ihre-umsetzung.html?utm_source=google&utm_medium=cpc&utm_campaign=CROSSMEDIA_ai&utm_id=sea&utm_content=text&utm_term=euaiact1&gad_source=1&gclid=CjwKCAjw48-vBhBbEiwAzqrZV

Paschou, Vasiliki. *Bias bei künstlicher Intelligenz: Risiken und Lösungsansätze.*
<https://www.activemind.legal/de/guides/bias-ki/>, 2024

Peterjohann, Horst. *Informationssicherheit. Die Schutzziele Vertraulichkeit, Integrität und Verfügbarkeit sicherstellen.*
<https://www.peterjohann-consulting.de/informationssicherheit/>, 2023

PrivacyXperts. *KI-Verordnung (AI-Act): Diese neuen Regelungen zu KI-Systemen und Fristen müssen Sie ab sofort beachten.*
<https://www.datenschutz-praemien.de/praemien/KI-Verordnung.pdf>

Protiviti. *Whitepaper: Next-Gen Risk Management.*
<https://www.protiviti.com/US-en/insights/whitepaper-nextgen-risk-management>, 2022,
nicht mehr Online verfügbar

Poretschkin, Dr. Maximilian; Schmitz, Anna; Akila, Dr. Maram; Adilova, Linara; Becker, Dr. Daniel; Cremers, Prof. Dr. Armin B.; Hecker, Dr. Dirk; Houben, Dr. Sebastian; Mock, PD Dr. Michael; Rosenzweig, Julia; Sicking, Joachim; Schulz, Elena; Voss, Dr. Ang. *Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz - KI-Prüfkatalog.*
https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatalog/202107_KI-Pruefkatalog.pdf

PSW. *Künstliche Intelligenz & Cybersecurity: Fluch und Segen zugleich.*
[https://www.psw-group.de/blog/kuenstliche-intelligenz-fluch-und-segen/8281?mtm_campaign=\(dach%3Ade%20max%3Aai\)%20Brand%20%7C%20Sales&mtm_kwd=&mtm_source=cpc&gad_source=1&gclid=CjwKCAiA7t6sBhAiEiwAsaieYhkB-elxJ-3UpP9nLZA86zdFeLKNy3PC1wPO5T7dGkzZJzyFmlhtsBo](https://www.psw-group.de/blog/kuenstliche-intelligenz-fluch-und-segen/8281?mtm_campaign=(dach%3Ade%20max%3Aai)%20Brand%20%7C%20Sales&mtm_kwd=&mtm_source=cpc&gad_source=1&gclid=CjwKCAiA7t6sBhAiEiwAsaieYhkB-elxJ-3UpP9nLZA86zdFeLKNy3PC1wPO5T7dGkzZJzyFmlhtsBo), 2021

SAP. *Was ist künstliche Intelligenz?*
<https://www.sap.com/austria/products/artificial-intelligence/what-is-artificial-intelligence.html>

Scribbr. *Scribbr-Plagiatsprüfung*
<https://app.scribbr.de>

Sinner, Fabian. *Warum Automatisierung und KI beim DDoS-Schutz wichtig sind.*
<https://www.link11.com/de/blog/it-sicherheit/automatisierung-ki-wichtig-fuer-ddos-schutz>, 2020

Shklyarov, Andrey; Vyrostkov, Dmitry. *Künstliche Intelligenz (KI) und maschinelles Lernen (ML) 7 Wege, wie KI und ML der Cybersicherheit helfen und schaden.*
<https://www.security-insider.de/7-wege-wie-ki-und-ml-der-cybersicherheit-helfen-und-schaden-a-059d2f88c3154b84dad8c3fe25c80ae/>, 2022

SOC-COMM: *SOC-COMM downloads.*
<https://www.soc-cmm.com/products/soc-cmm/> ff.

Splunk. *Splunk® Enterprise – Funktionen*
https://www.splunk.com/de_de/products/splunk-enterprise-features.html

Splunk. *A Beginner's Guide to Observability.*
<https://www.splunk.com/pdfs/ebooks/beginners-guide-to-observability.pdf>

Splunk. *Splunk® Enterprise – Funktionen.*
https://www.splunk.com/de_de/products/splunk-enterprise-features.html

Splunk. *A Beginner's Guide to Observability. Cutting through the complexity to learn what your systems, services and apps are really doing.*
<https://www.splunk.com/pdfs/ebooks/beginners-guide-to-observability.pdf>, 2021

Splunk. *Splunk Observability.*
https://www.splunk.com/en_us/products/observability.html#end-to-end-visibility

Splunk. *Splunk AI: Use-Case-Leitfaden für Einsteiger - Künstliche Intelligenz von Splunk für Observability.*
https://www.splunk.com/de_de/pdfs/gated/ebooks/splunk-machine-learning-for-observability-use-case-guide.pdf, 2023

Splunk. *TURN DATA INTO DOING Splunk Observability Cloud Free Trial.*
https://www.splunk.com/en_us/download/o11y-cloud-free-trial.html

SolarWinds. *Observability done right. Finally.*
<https://www.solarwinds.com/de/solarwinds-platform#>

SolarWinds. *Whitepaper - The Complete Guide to Keeping IT Simple.*
https://assets.contentstack.io/v3/assets/blt28ff6c4a2cf43126/blteb691741dc3f4134/65d4caa1c4151c79de122500/2401_ITSM_whitepaper_KeepITSimple.pdf, 2024

SolarWinds. *SolarWinds Predicts Key Trends and Themes That Will Define Enterprise IT in 2024.*
<https://investors.solarwinds.com/news/news-details/2023/SolarWinds-Predicts-Key-Trends-and-Themes-That-Will-Define-Enterprise-IT-in-2024/default.aspx>, 2023

SoSafe. *Cybercrime- Trends 2023.*
<https://sosafe-awareness.com/resources/reports/cybercrime-trends-2023/>, 2023

statworx GmbH. *AI ACT QUICK CHECK.*
https://www.statworx.com/ai-act/tool/Ergebnis_AI-Act-Quick-Check_nicht-betroffen.pdf

Spyder.
<https://www.spyder-ide.org/>

Syss.GmbH
<https://www.syss.de/>

Trabold, Dr. Daniel. *Welche Arten von Maschinellem Lernen gibt es?*
<https://lamarr-institute.org/de/blog/welche-arten-von-maschinellem-lernen-gibt-es/>, 2021

Wallner, Anna. *Neuronale Netze.*
https://www.mathematik.uni-ulm.de/stochastik/lehre/ss07/seminar_sl/ausarbeitung_wallner.pdf, 2007

Welke, Pascal. *Was ist eine lineare Regression?*
<https://websites.fraunhofer.de/ML-Blog/grundlagen/was-ist-eine-lineare-regression/>

WOLF, ARCTIC. *THE MARKET LEADER IN SECURITY OPERATIONS.*
<https://arcticwolf.com/resource/aw/siem-a-comprehensive-guide>, 2023

Wuttke, Vinzent. *Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion.*
<https://datasolut.com/neuronale-netzwerke-einfuehrung/>, 2024

Wuttke, Vinzent. *Machine Learning: Definition, Algorithmen, Methoden und Beispiele.*
<https://datasolut.com/was-ist-machine-learning/#unueberwachtes-lernen>, 2024

XGBoost dmlc. *XGBoost Documentation.*
<https://xgboost.readthedocs.io/en/latest/>

Gerda Žigienė. *Artificial Intelligence Based Commercial Risk Management Framework for SMEs.*
<https://www.mdpi.com/2071-1050/11/16/4501>, 2019

Zillmann, Mario; Partner, Lünendonk & Hossenfelder GmbH. *Cyber Security ist das Top-Thema für CIOs.*
<https://www.luenendonk.de/aktuelles/presseinformationen/cyber-security-ist-das-top-thema-fuer-cios/>, 2021

Zillmann, Mario; Partner, Lünendonk & Hossenfelder GmbH. *Cyber Security. Die digitale Transformation sicher gestalten.*

<https://s3.eu-central-1.amazonaws.com/cdn.a3bau.at/public/2021-02/Whitepaper-Luenendonk-ArvatoSystems-CyberSecurity.pdf>, 2020

Anhang VIII Hilfsmittelverzeichnis

Languagetool nur für Rechtschreib- und Grammatikkorrektur verwendet

https://languagetool.org/de/rechtschreibpruefung-deutsch?utm_source=google&utm_medium=cpc&utm_campaign=GA_LT_DACH_de_Sales_KW_BroadMatch_3Upgrade&utm_content=rechtschreibpr%C3%BCfung&utm_term=rechtschreibpr%C3%BCfung&gad_source=1&gclid=CjwKCAjwgfm3BhBeEiwAFfxrG6n346AJfZozuXQZPXLG_mE_JfgO2DWcgtnQmpjtnyPR8NqvJ4omexoCx8YQAvD_BwE

ChatGPT nur für Rechtschreib- und Grammatikkorrektur verwendet

<https://chatgpt.com>

Python-Programmierung mit Jupyter Notebooks für Analyse im Abschnitt 3.8 Risikoanalyse und 3.9 Erstellung von gesetzlichen Anforderungen nach BSI-Schema sowie im Anhang I und Anhang II

<https://jupyter.org>

Python-Programmierung mit Spyder für Analyse im Abschnitt 3.8 Risikoanalyse und 3.9 Erstellung von gesetzlichen Anforderungen nach BSI-Schema sowie im Anhang I und Anhang II

<https://www.spyder-ide.org/>

Plagiatsprüfung inkl. Quellen mit Scribbr

<https://app.scribbr.de>

Plagiatsprüfung inkl. KI-Erkennung mit GoThesis

<https://www.gothesis.de/>

Anlage IX Eidesstattliche Erklärung

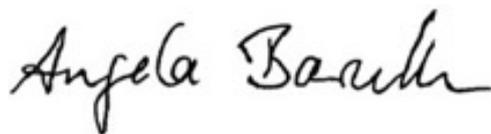
Name: Angela, Baruth

Matrikelnummer: q4470389

Hiermit versichere ich an Eides statt, dass ich diese Masterarbeit eigenständig verfasst und ausschließlich die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Passagen, die wörtlich oder sinngemäß aus Publikationen oder Vorträgen anderer Autoren übernommen wurden, habe ich ordnungsgemäß als solche gekennzeichnet.

Ich bin mit einer Plagiatsprüfung einverstanden.

Diese Masterarbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.



Neubrandenburg, 31.12.2024

Ort, Abgabedatum

Unterschrift (Vor- und Zuname)