Holger Reibold

Lokale KI unter Kontrolle

Sichere Architektur, Betrieb und Governance von GenAI- und RAG-Systemen

BRAIN-MEDIA.DE

Holger Reibold

Lokale KI unter Kontrolle

Sichere Architektur, Betrieb und Governance von GenAI- und RAG-Systemen

BRAIN-MEDIA.DE

ISBN: 978-3-95444-314-7

Cover: Freepik

Brain-Media.de

Dr. Holger Reibold – Hubert-Müller-Str. 52c – 66113 Saarbrücken

info@brain-media.de – www.brain-media.de

Inhaltsverzeichnis

Vorwort

Generative KI hat sich in erstaunlich kurzer Zeit von einem experimentellen Forschungsthema zu einer produktiven Infrastrukturkomponente entwickelt. Große Sprachmodelle beantworten Supportanfragen, analysieren interne Dokumente, unterstützen Entwickler und automatisieren Wissensarbeit. Parallel dazu wächst in vielen Organisationen der Wunsch, diese Fähigkeiten nicht ausschließlich über externe Cloud-Dienste zu nutzen, sondern unter eigener Kontrolle zu betreiben. Lokale oder privat betriebene GenAI-Plattformen versprechen Datenhoheit, geringere Latenzen und bessere Integrationsmöglichkeiten in bestehende IT-Landschaften. Doch mit der technischen Kontrolle wandert auch die volle Verantwortung für Sicherheit, Stabilität und Governance ins eigene Haus.

Viele erste Implementierungen folgen noch einem prototypischen Muster: Modell laden, API starten, Chat-Frontend anbinden – und produktiv gehen. Was in der Evaluierungsphase funktioniert, offenbart im Dauerbetrieb jedoch schnell strukturelle Schwächen. Große Sprachmodelle bringen eine eigene Angriffsdynamik mit. Prompts können manipuliert werden, Retrieval-Bestände lassen sich vergiften, und ungeschützte Inferenzendpunkte eröffnen neue Missbrauchsszenarien. Gleichzeitig verbinden moderne GenAI-Stacks mehrere bislang getrennte Disziplinen: GPU-beschleunigte Inferenz, datengetriebene Retrieval-Pipelines, API-Gateways, Container-Orchestrierung

und oft noch unreife Entwickler-Workflows. Risiken entstehen dabei selten durch einen einzelnen gravierenden Fehler, sondern durch unsaubere Übergänge zwischen diesen Komponenten.

Hinzu kommt eine begriffliche Unschärfe. In der Forschung bezeichnet „Private AI" – ich spreche in diesem Buch von lokaler KI – häufig Verfahren wie homomorphe Verschlüsselung, Differential Privacy oder föderiertes Lernen. In der betrieblichen Praxis hingegen geht es meist um lokal oder in kontrollierten Umgebungen ausgeführte GenAI-Workloads. Dieses Buch folgt bewusst der zweiten Perspektive: Im Mittelpunkt steht der sichere Entwurf, die Härtung und der stabile Betrieb lokal kontrollierter GenAI- und RAG-Systeme. Kryptographische Spezialverfahren werden dort betrachtet, wo sie architektonisch relevant sind, nicht als alleinige Lösung für KI-Sicherheit.

Der Schwerpunkt liegt klar auf umsetzbaren Maßnahmen. Statt abstrakter Grundsatzdiskussionen behandelt das Buch konkrete Architekturentscheidungen, typische Fehlkonfigurationen und praxiserprobte Schutzmechanismen. Anhand einer durchgängigen Referenzarchitektur wird gezeigt, wie sich eine lokale KI-Plattform strukturiert aufbauen und absichern lässt – von der Netzwerksegmentierung über die Absicherung der RAG-Pipeline bis hin zu Guardrails, Monitoring und Incident-Runbooks. Ziel ist eine Umgebung, die nicht nur funktional überzeugt, sondern auch unter Sicherheits-, Compliance- und Betriebsaspekten beherrschbar bleibt.

Das Buch richtet sich an Architekten, Plattform- und DevOps-Teams, Security Engineers sowie technisch orientierte Entscheider, die GenAI-Workloads im eigenen Verantwortungsbereich betreiben oder dies planen. Vorausgesetzt werden Grundkenntnisse in Container- und Cloud-nativen Umgebungen; tiefes Vorwissen zu großen Sprachmodellen ist hilfreich, aber nicht zwingend erforderlich.

Lokale KI wird in den kommenden Jahren fester Bestandteil moderner Unternehmensinfrastrukturen werden. Organisationen, die früh eine saubere Architekturdisziplin, klare Trust Boundaries und belastbare Betriebsroutinen etablieren, verschaffen sich einen nachhaltigen Vorteil.

Dieses Buch soll dabei als technischer Leitfaden dienen — mit dem Anspruch, aus schnellen Proof-of-Concepts dauerhaft kontrollierbare, sichere KI-Plattformen zu machen.

Ich wünsche Ihnen viel Erfolg beim Betrieb und der Absicherung lokaler KI-Systeme.

Herzlichst

Holger Reibold

1 Private KI im Infrastrukturkontext

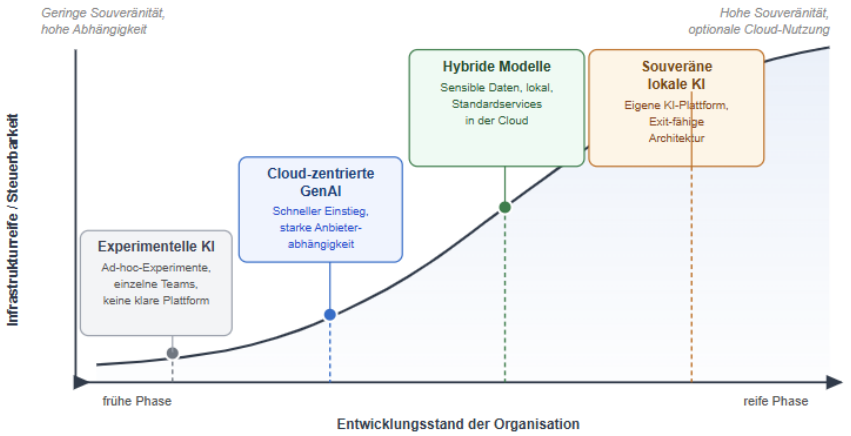
Lokale GenAI ist dort alternativlos, wo Datenabfluss regulatorisch oder organisatorisch nicht tolerierbar ist.

Lokale GenAI-Systeme entwickeln sich zunehmend von isolierten Pilotprojekten zu festen Bestandteilen der Unternehmensinfrastruktur. Während frühe Experimente meist mit minimalem Betriebs- und Sicherheitsfokus durchgeführt wurden, erfordern produktive Deployments eine deutlich höhere architektonische Disziplin. Modelle, Retrieval-Pipelines und Orchestrierungsschichten müssen heute wie kritische Plattformkomponenten behandelt werden – mit klaren Trust Boundaries, reproduzierbaren Deployments und überprüfbaren Datenflüssen.

Dieses Kapitel ordnet lokale beziehungsweise privat betriebene KI in den Infrastrukturkontext moderner IT-Landschaften ein. Im Mittelpunkt stehen die Treiber für On-Prem-GenAI, typische Einsatzmuster sowie die zentralen Architekturentscheidungen zwischen Cloud- und lokalem Betrieb. Ziel ist es, eine belastbare Entscheidungsgrundlage zu schaffen, bevor in den folgenden Kapiteln die konkrete Referenzarchitektur, Härtungsmaßnahmen und Betriebsprozesse vertieft werden. Wer die infrastrukturellen Rahmenbedingungen richtig

bewertet, vermeidet viele der späteren Sicherheits- und Skalierungsprobleme bereits in der Entwurfsphase.

Die in nachstehender Abbildung skizzierte Entwicklung zeigt den typischen Reifungspfad von KI-Infrastrukturen in Organisationen. Viele Initiativen beginnen in einer experimentellen Phase, die durch schnelle Cloud-Nutzung, geringe Governance und hohe Flexibilität geprägt ist. In diesem Stadium steht die funktionale Machbarkeit im Vordergrund, während Aspekte wie Kostenstabilität, Datenhoheit oder Betriebsreife noch eine untergeordnete Rolle spielen.



Typischer Entwicklungspfad von KI-Infrastrukturen von experimentellen Cloud-Nutzungen hin zu souveränen, produktionsreifen Plattformen mit zunehmender Governance-, Kosten- und Kontrolltiefe.

Mit wachsender Nutzung und zunehmender geschäftlicher Relevanz verschiebt sich der Fokus jedoch deutlich. Organisationen professionalisieren ihre Plattformen, standardisieren Betriebsmodelle und beginnen, regulatorische sowie ökonomische Effekte systematisch zu bewerten. In dieser Phase entstehen häufig hybride Architekturen, die Cloud-Agilität mit ersten lokalen Komponenten kombinieren.

Die rechte Seite des Entwicklungspfads markiert schließlich die Phase souveräner KI-Plattformen. Hier stehen kontrollierte Datenflüsse, reproduzierbarer Betrieb, Kosten-transparenz und technische Durchsetzbarkeit von Sicherheitsanforderungen im Mittelpunkt. Lokale oder dediziert betriebene GenAI-Workloads sind in diesem Stadium keine Gegenbewegung zur Cloud, sondern Ausdruck architektonischer Reife und bewusster Workload-Platzierung.

Hier zeigt sich: Der Übergang zu lokaler oder hybrider GenAI ist in vielen Organisationen kein disruptiver Strategiewechsel, sondern die konsequente Weiterentwicklung wachsender KI-Produktionsumgebungen.

1.1 Vom Experiment zur produktiven Plattform

Der Einsatz generativer KI beginnt in den meisten Organisationen experimentell. Einzelne Teams evaluieren Modelle, testen Chat-Oberflächen oder integrieren erste Retrieval-Funktionen in bestehende Anwendungen. Diese Phase ist bewusst explorativ: Geschwindigkeit zählt mehr als Robustheit, Sicherheitskontrollen sind minimal, und

die Infrastruktur ist häufig ad hoc aufgebaut. Typisch sind lokal gestartete Modellserver, einfache API-Gateways und unstrukturierte Dokumentenbestände für RAG-Tests.

Mit wachsender Nutzung verschieben sich jedoch die Anforderungen grundlegend. Sobald GenAI-Funktionen in produktive Workflows eingebunden werden – etwa im Support, in der Softwareentwicklung oder in internen Wissenssystemen – entstehen klassische Plattformanforderungen: Verfügbarkeit, Skalierbarkeit, Mandantentrennung, Nachvollziehbarkeit und Betriebssicherheit. Ein Sprachmodell ist dann nicht mehr nur ein Werkzeug, sondern Bestandteil einer geschäftskritischen Verarbeitungskette.

Der zentrale Fehler vieler Organisationen besteht darin, Proof-of-Concept-Architekturen nahezu unverändert in den Produktivbetrieb zu überführen. Dabei fehlen meist mehrere Schichten, die für einen stabilen Plattformbetrieb erforderlich sind: reproduzierbare Build- und Deployment-Pipelines, versionierte Modellartefakte, klar definierte Service-Identitäten, kontrollierte Netzwerkzonen und belastbares Monitoring. Besonders kritisch ist die Datenebene. In frühen Tests werden Dokumente häufig ungeprüft ingestiert und global durchsuchbar gemacht. Unter Produktionsbedingungen führt dies schnell zu Compliance- und Sicherheitsproblemen.

Ein weiterer Unterschied zwischen Experiment und Plattformbetrieb liegt in der Deterministik. Klassische Microservices liefern bei gleichen Eingaben reproduzierbare Ergebnisse. Große Sprachmodelle hingegen erzeugen probabilistische Antworten, deren Qualität stark

vom Prompt, vom Kontext und von der Modellversion abhängt. Daraus folgt, dass Konfigurationsänderungen – etwa neue Embeddings, geänderte Chunk-Größen oder aktualisierte Modellgewichte – unmittelbare Auswirkungen auf das Laufzeitverhalten haben können. Produktive Plattformen benötigen deshalb ein sauberes Versionierungs- und Rollback-Modell nicht nur für Code, sondern auch für Modelle und RAG-Datenbestände.

Auch die Bedrohungslage verändert sich mit zunehmender Exposition. Während interne Tests meist von vertrauenswürdigen Nutzern durchgeführt werden, sind produktive Systeme systematisch automatisierten Abfragen, Prompt-Injection-Versuchen und möglichen Datenabfluss-Szenarien ausgesetzt. Ohne Rate Limits, Eingabevalidierung, kontextbewusste Prompt-Trennung und mehrstufige Guardrails steigt das Risiko exponentiell mit der Nutzung.

Der Übergang zur produktiven GenAI-Plattform ist daher primär eine Frage der Architekturdisziplin. Erfolgreiche Umgebungen zeichnen sich durch klar getrennte Sicherheitszonen, identitätsbasierte Dienstkommunikation, streng kontrollierte Ingestion-Pipelines und durchgängige Observability aus. Erst wenn diese Grundlagen etabliert sind, lässt sich GenAI unter realer Last zuverlässig betreiben. Organisationen, die diesen Reifegrad systematisch erreichen, verwandeln experimentelle KI-Funktionen in eine dauerhaft beherrschbare Infrastrukturkomponente.

1.2 Treiber für On-Prem-GenAI

Der Trend zu lokal betriebenen GenAI-Systemen ist kein Selbstzweck, sondern die Folge konkreter technischer und regulatorischer Anforderungen. Drei Faktoren dominieren die Architekturentscheidungen in der Praxis: Datenkontrolle, Latenzverhalten und Compliance-Fähigkeit. Ihre Gewichtung variiert je nach Branche, bestimmt jedoch maßgeblich, ob Cloud-basierte LLM-Dienste ausreichen oder eine On-Prem-Plattform notwendig wird.

Der wichtigste Treiber ist die Datenhoheit. Viele produktive GenAI-Anwendungen greifen auf interne Wissensbestände zu – Quellcode, Verträge, Forschungsdaten, Supporttickets oder personenbezogene Informationen. In solchen Szenarien reicht eine vertragliche Zusicherung externer Anbieter oft nicht aus. Organisationen benötigen technisch überprüfbare Garantien darüber, wo Daten verarbeitet werden, welche Telemetrie abfließt und ob Inhalte in Trainingspipelines einfließen können. Lokale Deployments ermöglichen eine klare Kontrolle der Datenpfade, insbesondere wenn Modellendpunkte in isolierten Netzwerkzonen betrieben und ausgehende Verbindungen restriktiv gehandhabt werden. Für sicherheitskritische Umgebungen ist diese physische und logische Kontrollierbarkeit ein entscheidender Vorteil.

Die in nachstehender Abbildung dargestellten Einflussfaktoren wirken in der Praxis selten isoliert, sondern verstärken sich häufig gegenseitig. In vielen Organisationen ist beispielsweise nicht allein der

Datenschutz ausschlaggebend, sondern die Kombination aus regulatorischem Druck, steigenden Inferenzkosten und wachsenden Anforderungen an Betriebs-kontrolle. Das Kräftebild verdeutlicht damit, dass die Entscheidung für On-Prem-GenAI typischerweise aus einem mehrdimensionalen Bewertungsprozess hervorgeht.

Die zentralen Treiber für den Einsatz von On-Prem-GenAI. Die relative Gewichtung von Datenschutz, Latenz, Kosten, Kontrolle und Compliance bestimmt die optimale Platzierung von GenAI-Workloads.

Auffällig ist zudem, dass die Gewichtung der einzelnen Treiber stark vom Anwendungskontext abhängt. Während in hochregulierten Branchen Datenschutz und Compliance dominieren, stehen in interaktiven oder produktionsnahen Szenarien häufig Latenz und deterministische Performance im Vordergrund. Kosten entfalten ihre Wirkung oft zeitverzögert: Mit wachsendem Inferenzvolumen verschiebt sich die Wirtschaftlichkeitsbetrachtung nicht selten zugunsten lokaler Plattformen.

Der zweite wesentliche Faktor ist die Latenz. GenAI-Workloads, insbesondere RAG-Systeme mit großen Kontextfenstern, generieren hohe Datenvolumina pro Anfrage. Jede externe Round-Trip-Kommunikation erhöht die Antwortzeit und verschlechtert die Interaktivität. In Entwicklungsumgebungen, internen Assistenten oder automatisierten Workflows kann bereits eine zusätzliche Sekunde spürbare Produktivitätseinbußen verursachen. Lokale Inferenz reduziert Netzwerkabhängigkeiten und stabilisiert Antwortzeiten auch bei hoher Parallelität. Dieser Effekt wird mit zunehmender Tokenzahl, komplexeren Retrieval-Schritten und multimodalen Modellen weiter verstärkt.

Der dritte Treiber ist regulatorischer Druck. Datenschutzgesetze, branchenspezifische Vorgaben und interne Governance-Richtlinien verlangen nachvollziehbare Datenverarbeitung, feingranulare Zugriffskontrollen und auditierbare Systemzustände. Lokale GenAI-Plattformen erleichtern die Umsetzung dieser Anforderungen erheblich. Dokumentbasierte Zugriffskontrollen im RAG-Store,

Literatur- und Quellenverzeichnis

Abadi, M., et al. (2016). Deep Learning with Differential Privacy.

Advances in Neural Information Processing Systems, 35, 30318–30332. https://arxiv.org/abs/2208.07339

AMD (2025). The Rise of AI PCs in the Commercial Market: Planning for the Future. https://www.amd.com/en/blogs/2025/the-rise-of-ai-pcs-in-the-commercial-market-plann.html

Anthropic. (2022). Constitutional AI: Harmlessness from AI feedback. Online: https://www.anthropic.com.

Apple Inc. (2023). MLX: Efficient and flexible machine learning on Apple.

Arize AI. (2023). Phoenix: Open-source AI observability for LLMs and RAG.

Barnett, S., et al. (2024). Seven Failure Points When Engineering a Retrieval Augmented Generation System.

Cloud Native Computing Foundation. (2023). Kubernetes security best practices. https://kubernetes.io

Dettmers, T., et al. (2023). QLoRA: Efficient Finetuning of Quantized LLMs.

Dettmers, T., Lewis, M., Belkada, Y., & Zettlemoyer, L. (2022).

Dubey, A., et al. (2024). The Llama 3 herd of models. arXiv. https://arxiv.org/abs/2407.21783

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Online. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act). Online:

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

Frantar, E., Ashkboos, S., Hoefler, T., & Alistarh, D. (2023). GPTQ: Accurate post-training quantization for generative pre-trained transformers. International Conference on Learning Representations.

Gerganov, G. (2023). llama.cpp [Computer software]. GitHub. https://github.com/ggerganov/llama.cpp

Google. (2023). Secure AI Framework (SAIF). Google Cloud. https://cloud.google.com/security/saif

Han, S., Mao, H., & Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding.

Hu, E. J., et al. (2021). LoRA: Low-Rank Adaptation of Large Language Models.

Inan, H., et al. (2023). Llama Guard: LLM-based Input-Output Safeguard for Human-AI Conversations.

Jiang, A. Q., et al. (2023). Mistral 7B. arXiv:2310.06825.

Johnson, J., et al. (2019). Billion-scale similarity search with GPUs.

Khowaja, S. A., Dev, K., Qureshi, N. M. F., Khuwaja, P., & Foschini, L. (2022). Toward industrial private AI: A two-tier framework for data and model security. IEEE Wireless Communications, 29(2), 76–83.

Kwon, W., et al. (2023). Efficient memory management for large language model serving with PagedAttention. Proceedings of the ACM Symposium on Operating Systems Principles, 611–626.

Lauter, K. (2022). Private AI: machine learning on encrypted data. In Recent advances in industrial and applied mathematics (pp. 97-113). Cham: Springer International Publishing.

Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural

Information Processing Systems, 33, 9459–9474. https://arxiv.org/abs/2005.11401

Lin, J., Tang, J., Tang, H., Yang, S., Dang, X., & Han, S. (2023). AWQ: Activation-aware weight quantization for LLM compression and acceleration. arXiv. https://arxiv.org/abs/2306.00978

LocalAI Project. (2023). LocalAI: The free, Open Source OpenAI alternative [Computer software]. https://localai.io

McMahan, B., et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data.

Microsoft Corporation. (2021). ONNX Runtime. Online: https://onnxruntime.ai.

MITRE. (2023). Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS). https://atlas.mitre.org

NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://www.nist.gov

NVIDIA Corporation. (2023). TensorRT-LLM. Online: https://github.com/NVIDIA/TensorRT-LLM

Ollama Team. (2023). Ollama [Computer software]. https://ollama.com

Open Source Initiative (OSI). (2024). The Open Source AI Definition.

OWASP Foundation. (2023). OWASP Top 10 for large language model applications. https://owasp.org.

Rebedea, T., et al. (2023). NeMo Guardrails: A Toolkit for Steerable and Safe LLM Applications. (NVIDIA).

Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

von Platen, P., et al. (2022). Diffusers: State-of-the-art diffusion models. GitHub.

Weidinger, L., et al. (2021). Ethical and social risks of harm from language models.

Wolf, T., et al. (2020). Transformers: State-of-the-art natural language processing. EMNLP System Demonstrations, 38–45.

Xiao, G., Lin, J., Seznec, M., Wu, H., Demouth, J., & Han, S. (2023). SmoothQuant: Accurate and efficient post-training quantization for large language models. International Conference on Machine Learning.

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE, 107(8), 1738–1762.

Stichwortverzeichnis

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

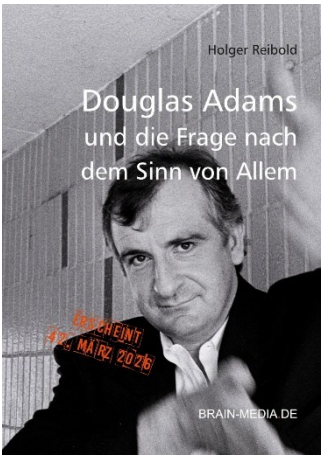
V

W

X

Z

Mehr von Brain-Media.de



42 – Douglas Adams und die Frage nach dem Sinn von Allem

Am 11. Mai 2026 ist Douglas Adams 25 Jahre tot. Der Kultautor hat der Welt wunderbar, skurrile Werke geschenkt. Jetzt ist es an der Zeit, den Autor kennenzulernen.

Umfang: 140 Seiten

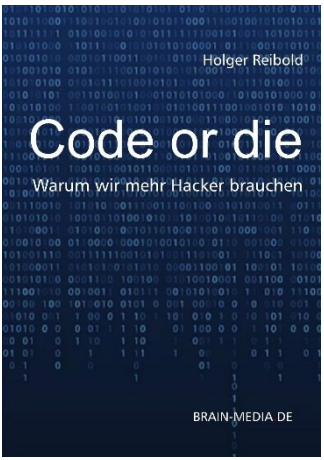
Preis: 14,99 EUR

Erscheint: 42. März 2026

Towelday, das ultimative Handtuch für alle Fans

An seinem Todestag, dem Towelday, erinnern sich Fans an Douglas Adams und huldigen dem Kultautor.

100 % intergalaktisch geprüfte Baumwolle, nachhaltig Produktion zum Preis von 42 EUR.



Code or die – Warum wir mehr Hacker brauchen

Ein Manifest für mehr digitale Selbstbestimmung, Neugierde und Eigenverantwortung. Medienkompetenzen alleine genügen nicht; die Gesellschaft von morgen braucht Digitalkompetenzen.

Umfang: 120 Seiten

Preis: 14,99 EUR

Erscheint Frühjahr 2026

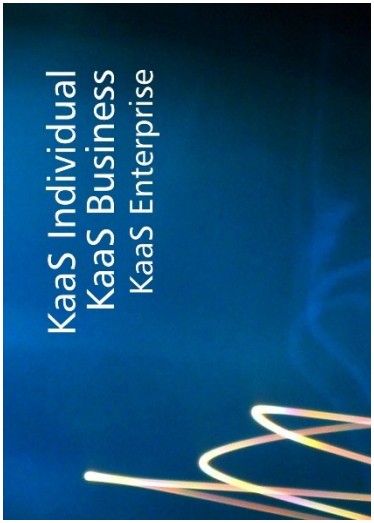
KI Incident Response – Wie man Sicherheitsvorfälle in KI-Systemen erkennt, eindämmt und verantwortet

Ziel- und punktgenaue Reaktionen für kritischen KI-Vorfälle.

Umfang: 140 Seiten

Preis: 16,99 EUR

Erscheint: Frühjahr 2026



Brain-Media KaaS

Individual

Business

Enterprise

– die E-Book-Flatrate –

Brain-Media.de ist ein **Pionier** in Sachen E-Book-Flatrates. Bereits 2014 wurden mit Plus+ E-Books als attraktives und preisgünstiges Abo angeboten. Im April 2026 erlebt Plus+ eine Renaissance.

Abonnementen erhalten **1 Jahr Zugriff auf alle E-Books**, die bereits erschienen sind, und auf alle Neuerscheinungen des Abozeitraums. Sie können über 30 verfügbare E-Books herunterladen – dazu editierbare Checklisten, Vorlagen, Beispiele etc. Sie können die E-Books ausdrucken und auf mehreren Lesegeräten verwenden. Die E-Books verwenden kein DRM!

Und das alles zum **unschlagbaren Sonderpreis**!

IT-Texter.one

100+ IT-Fachbücher

1500+ Fachartikel

30+ Erfahrung

KOMPLEXE INHALTE PUNKTGENAU AUFZUBEREITEN, IST EINE KUNST. ICH BEHERRSCHE SIE. BEI MIR ERHALTEN SIE FACH-TEXTE, DIE KOMPLEXES VERSTÄNDLICH MACHEN.

Seit über 30 Jahren unterstütze ich Unternehmen aus der IT-, Software- und Digitalbranche dabei, ihre technischen Inhalte klar, präzise und zielgruppenorientiert zu kommunizieren. Als promovierter Informatiker und erfahrener IT-Journalist verbinde ich fundiertes Fachwissen mit journalistischem Storytelling. Als Key Account Manager eines IT-Dienstleisters verfüge ich obendrein über konkrete Erfahrungen mit allen gängigen Technologien.

WARUM SIE MIT MIR ARBEITEN SOLLTEN

35 Jahre Erfahrung mit Internet-, Netzwerk- und Webtechnologien

Kooperation mit führenden Akteuren der IT- und Medienbranche

Strategisches Denken: Texte, die nicht nur informieren, sondern auch verkaufen

THEMENSCHWERPUNKTE	WIE KANN ICH SIE UNTERSTÜTZEN?
Open-Source	Content Creation
Enterprise IT	Dokumentationen
IT-Consulting	Case Studies
SaaS	Suchmaschinenoptimierung
Künstliche Intelligenz	Tech-Marketing

MEIN VERSPRECHEN

Ich übernehme die inhaltliche und sprachliche Brücke zwischen Technologie und Anwendung. Selbst komplexe Sachverhalte kommen beim Publikum an – fachlich korrekt, prägnant und SEO-wirksam.

PREISMODELLE

Professionelle Leistungen, die ihresgleichen suchen, gibt es nicht umsonst. Sprechen Sie mit an. Gerne vereinbaren wir einen Fixpreis; das vereinfacht Ihre Kalkulation.

KONTAKT AUFNEHMEN

Sprechen wir über Ihr Projekt. Schreiben Sie mir eine Mail (info@it-texter.one). Oder besser noch: Rufen Sie mich an (+49 681 91005698).