Holger Reibold

# Autonomous Red Teaming mit KI

Von manuellen Pentests zu autonomen Angreifern: KI-basierte Systeme für kontinuierliche und adaptive Sicherheitsüberprüfung

BRAIN-MEDIA.DE

Holger Reibold

# Autonomous Red Teaming mit KI

Von manuellen Pentests zu autonomen Angreifern: KI-basierte Systeme für kontinuierliche und adaptive Sicherheitsüberprüfung

BRAIN-MEDIA.DE

ISBN: 978-3-95444-347-5

Cover: Freepik

Brain-Media.de

Dr. Holger Reibold – Huber-Müller-Str. 52 – 66113 Saarbrücken

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

# Inhaltsverzeichnis

# Vorwort

*Autonome Sicherheit: Warum sich alles grundlegend ändert*

Die Sicherheit von IT-Systemen steht an einem Wendepunkt. Über Jahrzehnte hinweg bildeten manuelle Penetrationstests das Rückgrat der technischen Sicherheitsüberprüfung. Erfahrene Spezialisten simulierten Angriffe, identifizierten Schwachstellen und lieferten wertvolle Erkenntnisse zur Verbesserung der Sicherheitslage. Dieses Modell war erfolgreich – aber es war immer auch begrenzt: zeitlich, personell und methodisch. In einer Welt, in der sich Systeme kontinuierlich verändern und Angreifer automatisiert, skalierend und persistent agieren, stößt dieses klassische Vorgehen zunehmend an seine Grenzen.

Die Motivation für dieses Buch liegt genau in dieser Diskrepanz. Klassische Pentests sind Momentaufnahmen. Sie liefern punktuelle Einsichten in eine dynamische Realität. Zwischen zwei Tests können sich Architekturen verändern, neue Schwachstellen entstehen oder bestehende Sicherheitsmaßnahmen obsolet werden. Gleichzeitig agieren reale Angreifer nicht in festen Intervallen, sondern kontinuierlich. Sie automatisieren Reconnaissance, nutzen skalierbare Werkzeuge und passen ihre Strategien dynamisch an ihre Ziele an. Die Frage ist daher nicht mehr, ob ein System zu einem bestimmten Zeitpunkt sicher

ist, sondern ob es dauerhaft unter realistischen Angriffsbedingungen bestehen kann.

Parallel dazu hat sich die Bedrohungslandschaft fundamental gewandelt. Angriffe sind heute hochgradig organisiert, oft durch automatisierte Kampagnen unterstützt und zunehmend durch künstliche Intelligenz verstärkt. Angreifer nutzen legitime Werkzeuge, bewegen sich unauffällig innerhalb von Systemen und verfolgen adaptive Strategien, die sich an Kontext, Umgebung und Abwehrmechanismen orientieren. Der klassische Gegensatz von „Angriff" und „Verteidigung" wird dabei zunehmend durch ein dynamisches Zusammenspiel ersetzt, in dem Geschwindigkeit, Anpassungsfähigkeit und Skalierung entscheidend sind.

Vor diesem Hintergrund entsteht eine neue Generation der Sicherheitsüberprüfung: Autonomous Red Teaming mit KI. Statt punktueller Tests treten kontinuierliche, automatisierte und adaptive Angriffssimulationen in den Vordergrund. KI-gesteuerte Agenten agieren als autonome Angreifer, die Systeme nicht nur analysieren, sondern aktiv erkunden, Schwachstellen identifizieren und Angriffspfade eigenständig entwickeln. Sie lernen aus ihren Aktionen, passen ihre Strategien an und simulieren damit realitätsnäher als je zuvor das Verhalten moderner Bedrohungsakteure.

Autonomie ist dabei nicht nur ein technologischer Fortschritt, sondern ein konzeptioneller Paradigmenwechsel. Sicherheitsüberprüfung wird von einem projektbasierten Ansatz zu einem permanenten Prozess. Systeme werden nicht mehr nur getestet, sondern

kontinuierlich „herausgefordert“. Dabei entstehen neue Anforderungen an Architektur, Steuerung und Governance. Autonome Systeme müssen kontrollierbar, nachvollziehbar und sicher betrieben werden. Gleichzeitig eröffnen sie die Möglichkeit, Sicherheitslücken schneller zu erkennen, Risiken präziser zu bewerten und Abwehrmaßnahmen gezielter zu verbessern.

Dieses Buch richtet sich an eine Zielgruppe, die diesen Wandel aktiv gestalten will. Dazu gehören Sicherheitsarchitekten, Red- und Blue-Teams, DevSecOps-Verantwortliche sowie Entscheidungsträger auf Managementebene. Es richtet sich an alle, die verstehen wollen, wie sich Sicherheitsüberprüfung in einer KI-getriebenen Welt verändert – und wie sich diese Veränderung praktisch umsetzen lässt.

Der Anspruch dieses Buches ist es, nicht nur Konzepte zu beschreiben, sondern ein strukturiertes Verständnis für autonome Red-Teaming-Systeme zu vermitteln. Es zeigt, wie KI als Enabler genutzt werden kann, welche architektonischen Prinzipien zugrunde liegen und wie sich kontinuierliche Sicherheitsüberprüfung in bestehende Organisationen integrieren lässt. Dabei werden sowohl technische als auch strategische und regulatorische Aspekte berücksichtigt.

Autonomous Red Teaming ist kein fernes Zukunftsszenario. Es ist eine notwendige Antwort auf eine Realität, in der Angriffe bereits heute automatisiert, adaptiv und skalierbar sind. Wer Sicherheit weiterhin als statischen Zustand begreift, wird dieser Realität nicht gerecht. Sicherheit wird zu einem dynamischen Prozess – und dieses

Buch liefert die Grundlage, diesen Prozess zu verstehen und aktiv zu gestalten.

Die folgenden Kapitel führen Schritt für Schritt durch diesen Wandel: von den Grenzen klassischer Pentests über die Rolle von KI bis

Dabei wünsche Ihne Ihnen viel Vergnügen.

Herzlichst

Holger Reibold

# 1 Vom Pentest zur Continuous Security

*Vom Pentest zur kontinuierlichen Sicherheitsprüfung*

Die Sicherheitsüberprüfung von IT-Systemen befindet sich in einem fundamentalen Wandel. Während klassische Penetrationstests lange als Goldstandard galten, stoßen sie zunehmend an strukturelle Grenzen. Sie sind punktuell, ressourcenintensiv und bilden nur einen Momentzustand ab – in einer Realität, in der sich Systeme kontinuierlich verändern und Angreifer dauerhaft aktiv sind.

Moderne Bedrohungsakteure agieren automatisiert, skalierbar und adaptiv. Sie nutzen vorhandene Systemwerkzeuge, passen ihr Verhalten dynamisch an und verfolgen langfristige Strategien. Diese Entwicklung stellt traditionelle Sicherheitsmodelle vor eine zentrale Herausforderung: Wie kann Sicherheit bewertet werden, wenn sich sowohl Systeme als auch Angriffe permanent weiterentwickeln?

Autonomous Red Teaming mit KI bietet darauf eine neue Antwort. KI-gesteuerte Agenten übernehmen die Rolle autonomer Angreifer, die Systeme kontinuierlich analysieren, Schwachstellen identifizieren und Angriffspfade eigenständig entwickeln. Dabei entsteht eine Sicherheitsüberprüfung, die nicht nur realistischer, sondern auch skalierbarer und effizienter ist.

Dieses Buch zeigt, wie sich dieser Paradigmenwechsel technisch, organisatorisch und strategisch umsetzen lässt.

## 1.1 Grenzen klassischer Pentests

Klassische Penetrationstests gelten seit vielen Jahren als zentrales Instrument zur Bewertung der IT-Sicherheit. Sie liefern wertvolle Einblicke in Schwachstellen, Fehlkonfigurationen und potenzielle Angriffsvektoren. Dennoch basieren sie auf einem grundlegenden Paradigma, das in der heutigen, hochdynamischen IT-Landschaft zunehmend an Wirksamkeit verliert: dem Prinzip der punktuellen Prüfung.

Ein Pentest bildet stets nur einen definierten Zeitraum ab. Er ist projektbasiert, zeitlich begrenzt und folgt in der Regel einem klar abgegrenzten Scope. Diese Struktur ist aus organisatorischer Sicht sinnvoll, führt jedoch zu einem entscheidenden Nachteil: Die Ergebnisse sind bereits kurz nach Abschluss potenziell veraltet. In modernen Umgebungen, in denen Systeme kontinuierlich weiterentwickelt, Deployments automatisiert durchgeführt und Konfigurationen regelmäßig angepasst werden, entsteht eine Diskrepanz zwischen Testzeitpunkt und realem Betriebszustand.

Ein weiterer limitierender Faktor ist die Skalierbarkeit. Manuelle Tests sind stark von der Expertise einzelner Spezialisten abhängig. Hochqualifizierte Red Teams sind begrenzt verfügbar, kostenintensiv und können nur eine begrenzte Anzahl an Systemen und Szenarien gleichzeitig abdecken. Gleichzeitig wachsen IT-Landschaften in

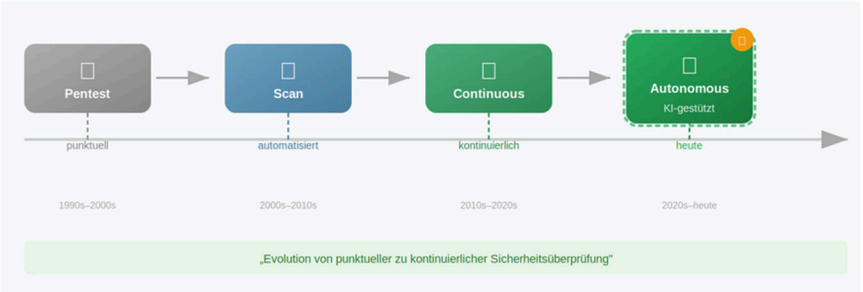
Umfang und Komplexität – insbesondere durch Cloud-Infrastrukturen, Microservices und hybride Architekturen. Die Folge ist eine strukturelle Unterdeckung: Nicht alle Systeme können in der notwendigen Tiefe und Frequenz geprüft werden.

Hinzu kommt die methodische Begrenzung klassischer Pentests. Obwohl erfahrene Tester kreativ und adaptiv agieren, bleiben ihre Aktivitäten innerhalb eines definierten Rahmens. Zeitdruck, Scope-Einschränkungen und operative Restriktionen führen dazu, dass nur ein Ausschnitt möglicher Angriffspfade untersucht wird. Komplexe, mehrstufige Angriffsszenarien – insbesondere solche, die über längere Zeiträume hinweg entstehen – bleiben häufig unentdeckt.

Ein besonders kritischer Punkt ist die fehlende Kontinuität. Reale Angreifer operieren nicht in Projektzyklen. Sie beobachten Systeme über längere Zeiträume, sammeln Informationen, testen Hypothesen und passen ihre Strategien iterativ an. Klassische Pentests hingegen sind diskrete Ereignisse. Zwischen zwei Tests existiert oft eine Phase ohne aktive Angriffssimulation, in der neue Schwachstellen entstehen können, ohne erkannt zu werden.

Darüber hinaus ist die Realitätsnähe begrenzt. Viele Tests erfolgen unter kontrollierten Bedingungen, mit abgestimmten Zeitfenstern und klar definierten Regeln. Dies ist notwendig, um Risiken für den Betrieb zu minimieren, führt jedoch dazu, dass bestimmte Angriffstechniken – insbesondere solche mit potenziell disruptiven Auswirkungen – nicht vollständig simuliert werden. Das resultierende

Sicherheitsbild ist daher häufig konservativ und nicht vollständig repräsentativ für reale Bedrohungsszenarien.




**Die Sicherheitsüberprüfung hat sich von punktuellen Pentests hin zu kontinuierlichen, automatisierten und zunehmend autonomen Prozessen entwickelt, die dynamische IT-Umgebungen realitätsnah abbilden.**

Schließlich spielt auch die Dokumentation eine Rolle. Pentest-Berichte sind in der Regel statische Artefakte, die den Zustand zum Testzeitpunkt widerspiegeln. Sie liefern Handlungsempfehlungen, jedoch keine kontinuierliche Rückkopplung darüber, ob identifizierte Schwachstellen tatsächlich behoben wurden oder ob neue Risiken entstanden sind. Die Verbindung zwischen Analyse und operativer Sicherheitssteuerung bleibt damit oft fragmentiert.

Zusammenfassend lässt sich festhalten: Klassische Pentests sind weiterhin ein wichtiges Werkzeug, aber sie sind nicht mehr ausreichend, um die Sicherheit moderner IT-Systeme umfassend und

nachhaltig zu bewerten. Ihre zeitliche Begrenzung, eingeschränkte Skalierbarkeit, methodische Fokussierung und fehlende Kontinuität machen sie zu einem Baustein – jedoch nicht zur alleinigen Lösung.

Diese Grenzen bilden den Ausgangspunkt für die Entwicklung neuer Ansätze. Autonomous Red Teaming mit KI adressiert genau diese Schwächen, indem es Sicherheitsüberprüfung von einem punktuellen Ereignis zu einem kontinuierlichen, adaptiven Prozess transformiert.

## 1.2 Realität moderner Angreifer

Die Bedrohungslandschaft im Cyberraum hat sich in den vergangenen Jahren grundlegend verändert. Angreifer agieren heute nicht mehr isoliert, opportunistisch oder ausschließlich manuell, sondern zunehmend organisiert, automatisiert und strategisch. Moderne Angriffe sind das Ergebnis strukturierter Prozesse, die auf Skalierbarkeit, Persistenz und Anpassungsfähigkeit ausgelegt sind.

Ein zentrales Merkmal moderner Angreifer ist ihre Fähigkeit zur Automatisierung. Reconnaissance, also die Informationsbeschaffung über Zielsysteme, erfolgt heute häufig kontinuierlich und ohne direkte menschliche Interaktion. Öffentliche Schnittstellen, Konfigurationsfehler oder neu exponierte Dienste werden automatisiert erkannt und bewertet. Dadurch entsteht ein permanenter „Scan-Zustand“, in dem potenzielle Angriffsflächen nahezu in Echtzeit identifiziert werden.

Darüber hinaus zeichnen sich moderne Angriffe durch ihre Persistenz aus. Angreifer verfolgen langfristige Strategien, die nicht auf schnellen Erfolg ausgerichtet sind, sondern auf nachhaltigen Zugriff. Initiale Kompromittierungen dienen oft nur als Einstiegspunkt, um sich schrittweise innerhalb eines Systems oder Netzwerks zu bewegen. Laterale Bewegung, Privilegieneskalation und das Ausnutzen vertrauensbasierter Beziehungen sind zentrale Bestandteile dieses Vorgehens.

Ein weiterer entscheidender Faktor ist die Adaptivität. Moderne Angreifer reagieren dynamisch auf die jeweilige Zielumgebung. Sie passen ihre Techniken an vorhandene Sicherheitsmechanismen an, vermeiden auffällige Aktivitäten und nutzen bevorzugt legitime Systemfunktionen. Dieses Vorgehen wird häufig als „Living off the Land" bezeichnet: Statt eigene Schadsoftware einzuschleusen, bedienen sich Angreifer vorhandener Werkzeuge wie PowerShell, administrativer Schnittstellen oder Standardprotokolle. Dadurch reduzieren sie ihre Sichtbarkeit und erhöhen die Erfolgswahrscheinlichkeit.

Auch die Skalierung spielt eine zentrale Rolle. Durch den Einsatz automatisierter Werkzeuge und Plattformen können Angriffe parallel gegen eine Vielzahl von Zielen durchgeführt werden. Cloud-Infrastrukturen, APIs und standardisierte Technologien bieten hierfür ideale Voraussetzungen. Angreifer sind in der Lage, erfolgreiche Muster schnell zu replizieren und auf neue Ziele anzuwenden.

Nicht zuletzt wird künstliche Intelligenz zunehmend selbst zum Werkzeug von Angreifern. KI-gestützte Systeme können Muster erkennen,

Entscheidungen unterstützen und Angriffspfade optimieren. Sie ermöglichen es, komplexe Zusammenhänge schneller zu analysieren und dynamisch auf Veränderungen zu reagieren. Damit verschiebt sich das Gleichgewicht weiter zugunsten adaptiver, lernender Angriffsstrategien.

Diese Entwicklung führt zu einer grundlegenden Herausforderung für klassische Sicherheitsansätze. Wenn Angriffe kontinuierlich, automatisiert und adaptiv erfolgen, reicht eine punktuelle Sicherheitsüberprüfung nicht mehr aus. Die Realität moderner Angreifer erfordert ein ebenso dynamisches Gegenmodell – eines, das kontinuierlich testet, sich anpasst und realistische Angriffsszenarien abbildet.

Genau hier setzt Autonomous Red Teaming mit KI an: Es überträgt die Prinzipien moderner Angreifer in kontrollierte, systematische Sicherheitsüberprüfungen und schafft damit eine neue Grundlage für realitätsnahe und nachhaltige Sicherheitsbewertung.

## 1.3 Sicherheitsüberprüfung

Die zunehmende Dynamik moderner IT-Landschaften stellt traditionelle Modelle der Sicherheitsüberprüfung grundlegend infrage. Systeme sind heute nicht mehr statisch, sondern unterliegen einem permanenten Wandel. Neue Funktionen werden kontinuierlich entwickelt, Deployments erfolgen automatisiert über CI/CD-Pipelines, und Infrastruktur wird flexibel in Cloud-Umgebungen skaliert. In diesem

Kontext verliert die Vorstellung, Sicherheit in festen Intervallen überprüfen zu können, ihre Gültigkeit.

Sicherheitsüberprüfung muss daher als kontinuierlicher Prozess verstanden werden. An die Stelle punktueller Tests tritt ein dauerhaftes, systematisches Beobachten, Analysieren und Herausfordern von IT-Systemen. Ziel ist es nicht mehr, einen „sicheren Zustand" zu einem bestimmten Zeitpunkt zu bestätigen, sondern die Fähigkeit eines Systems zu bewerten, sich unter realistischen Angriffsbedingungen dauerhaft zu behaupten.

Ein kontinuierlicher Ansatz basiert auf mehreren zentralen Prinzipien. Erstens: Permanenz. Systeme werden nicht sporadisch geprüft, sondern laufend überwacht und getestet. Neue Angriffsflächen, die durch Änderungen in der Infrastruktur entstehen, werden zeitnah erkannt und bewertet. Zweitens: Automatisierung. Die Vielzahl möglicher Angriffsszenarien und die Geschwindigkeit von Systemveränderungen erfordern skalierbare Mechanismen, die ohne permanente manuelle Eingriffe funktionieren. Drittens: Adaptivität. Sicherheitsüberprüfung muss in der Lage sein, sich an unterschiedliche Umgebungen, Konfigurationen und Bedrohungslagen anzupassen.

Ein weiterer wesentlicher Aspekt ist die Integration in bestehende Prozesse. Kontinuierliche Sicherheitsüberprüfung ist kein isoliertes Instrument, sondern Teil eines umfassenden Sicherheits- und Entwicklungsmodells. Sie muss eng mit DevSecOps-Praktiken, Monitoring-Systemen und Incident-Response-Prozessen verzahnt sein. Nur

so können erkannte Schwachstellen schnell bewertet, priorisiert und behoben werden.

Darüber hinaus verändert sich auch die Art der Ergebnisse. Statt statischer Berichte entstehen dynamische Feedback-Schleifen. Sicherheitsinformationen werden in Echtzeit bereitgestellt, kontinuierlich aktualisiert und direkt in operative Entscheidungen eingebunden. Dies ermöglicht eine deutlich engere Kopplung zwischen Analyse und Handlung und erhöht die Reaktionsgeschwindigkeit gegenüber neuen Risiken.

Ein kontinuierlicher Ansatz bringt jedoch auch neue Herausforderungen mit sich. Die Menge an generierten Daten steigt erheblich, was Anforderungen an Auswertung, Priorisierung und Visualisierung stellt. Gleichzeitig müssen Mechanismen etabliert werden, um Fehlalarme zu reduzieren und die Relevanz von Erkenntnissen sicherzustellen. Nicht zuletzt erfordert der Betrieb kontinuierlicher Systeme klare Governance-Strukturen, um Risiken zu kontrollieren und die Stabilität produktiver Umgebungen nicht zu gefährden.

Trotz dieser Herausforderungen ist der Übergang zu kontinuierlicher Sicherheitsüberprüfung unausweichlich. Die Geschwindigkeit und Komplexität moderner IT-Systeme sowie die Arbeitsweise heutiger Angreifer lassen keine Alternative zu. Sicherheit wird damit von einer punktuellen Prüfung zu einem integralen Bestandteil des laufenden Systembetriebs.

Autonomous Red Teaming mit KI stellt in diesem Kontext einen entscheidenden Baustein dar. Es ermöglicht, kontinuierliche Sicherheitsüberprüfung nicht nur skalierbar umzusetzen, sondern auch realitätsnah zu gestalten – indem es die Dynamik und Adaptivität moderner Angriffe in den Prüfprozess integriert.

## 1.4 Adaptive Sicherheitsmodelle

Mit der Transformation von IT-Systemen hin zu dynamischen, verteilten und hochgradig automatisierten Architekturen steigen auch die Anforderungen an Sicherheitsmodelle. Klassische, statische Ansätze – die auf festen Regeln, periodischen Prüfungen und klar abgegrenzten Systemgrenzen basieren – reichen nicht mehr aus, um der Komplexität moderner Umgebungen gerecht zu werden. Stattdessen sind adaptive Sicherheitsmodelle erforderlich, die sich kontinuierlich an veränderte Bedingungen anpassen können.

Ein zentrales Merkmal adaptiver Sicherheitsmodelle ist ihre Kontextsensitivität. Sicherheit kann nicht mehr isoliert auf einzelne Systeme oder Komponenten betrachtet werden, sondern muss im Gesamtzusammenhang verstanden werden. Faktoren wie Benutzerverhalten, Systeminteraktionen, Netzwerkbeziehungen und externe Einflüsse spielen eine entscheidende Rolle. Adaptive Modelle müssen in der Lage sein, diese Kontexte zu erfassen und ihre Bewertungen entsprechend anzupassen.

# Stichwortverzeichnis

## H



## I



## J



## K



## L



## M



## N



## O

## P



## R



## S



## T

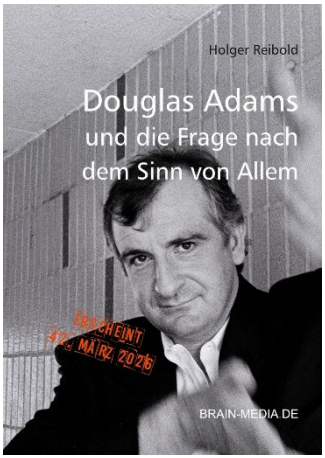


## V



## 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.


**Synergie der Intelligenz – Das Handbuch für das Design und die Implementierung von Multi-Agenten-Systemen**

Dieses Buch zeigt, wie Agenten zusammenarbeiten und wie Sie intelligente, skalierbare Systeme erfolgreich designen und einsetzen.

Umfang: 190 Seiten

Preis: 29,99 EUR


**Lokale KI – Sichere Architektur, Betrieb und Governance von GenAI- und RAG-Systemen**

RAG- und LLM-Plattformen mit klarer Architektur, Guardrails, Monitoring und Governance kontrolliert und resilient betreiben.

Umfang: 270 Seiten

Preis: 29,99 EUR


**Knowledge as a Service**

**Personal**

**Business**

**Enterprise**

IT-Security, Compliance und KI entwickeln sich schneller als jedes gedruckte Buch. Um dieser Dynamik Rechnung zu tragen, hat Brain-Media.de **KaaS – Knowledge as a Service** entwickelt.

Mit KaaS erhalten Sie ein lebendes **Wissenssystem**: Alle Titel als PDF/E-Book, **regelmäßig aktualisierte Living Documents** sowie **exklusive Downloads** – Checklisten, Vorlagen und sofort einsetzbare Templates. Alle Inhalte auch als **Audio – unterwegs nutzbar**.

Speziell für **Regulierung und Audits**: Inhalte zu NIS-2, DORA, CRA & AI Act werden laufend gepflegt und helfen Ihnen, An-forderungen strukturiert umzusetzen und auditfähig zu bleiben. Für fortgeschrittene Nutzung stehen Inhalte zusätzlich als **JSON, Markdown- und SCORM-Rohdaten** bereit – ideal für die Automatisierung und Integration in Ihre Umgebungen.

KaaS ist die wachsende **Bibliothek für**
**Praxis, Compliance und Resilienz**.