CLOUDFLARE

E-BOOK

Sicherer Umgang mit KI

Ein Leitfaden zur Entwicklung einer skalierbaren KI-Strategie für CISO

Inhalt

Kurzfassung

Herzlich Willkommen, CISO!

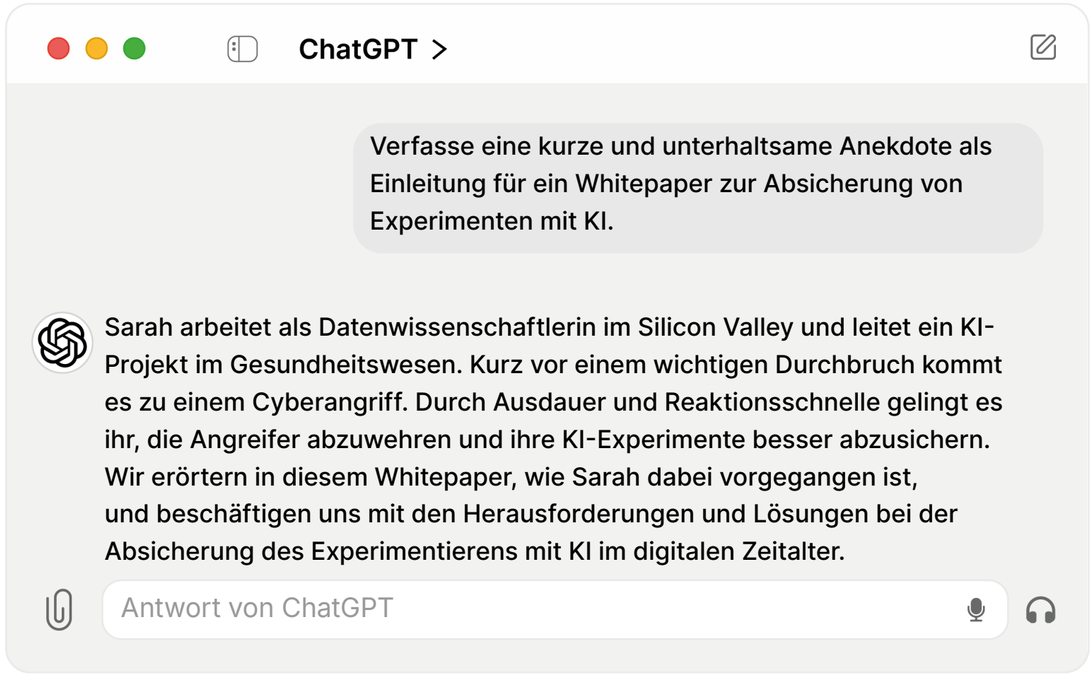
Kaum ein Schlagwort dürfte im öffentlichen Diskurs im Moment so oft auftauchen wie KI. Und da sie unter dem Sicherheitsaspekt eine große Herausforderung darstellt, kommt um das Thema heute niemand mehr herum. Der folgende Cloudflare-Leitfaden soll Ihnen bei Ihren Überlegungen helfen, wie Sie auf sichere Weise mit dem Einsatz von generativer künstlicher Intelligenz (Generative Artificial Intelligence – GenAI) in Ihrem Unternehmen experimentieren können.

KI wird nicht nur immer leistungsfähiger, sondern steht auch einer wachsenden Zahl von Anwendern zur Verfügung. Dadurch eröffnen sich in verschiedenen Branchen Innovationsmöglichkeiten. Doch wie andere große Umwälzungen bringt auch GenAI einzigartige Herausforderungen hinsichtlich Sicherheit, Datenschutz und Compliance mit sich. Mit ihrer zunehmenden Verbreitung können unerwartete Spitzen bei den Nutzerzahlen, Fälle von Missbrauch durch Anwender und das Entstehen gefährlicher Schatten-IT einhergehen. Dadurch steigt die Wahrscheinlichkeit, dass sensible Daten gestohlen werden oder an die Öffentlichkeit gelangen.

Doch diese Technologie erfreut sich in der Arbeitswelt einer immer größeren Beliebtheit. Deshalb benötigen Sie ein GenAI-Konzept, mit dem sie Produkte und Dienstleistungen maßstabsgerecht benutzen, entwickeln und schützen können. Dafür sollten Sie sich mit den damit verbundenen Risiken auseinandersetzen. Wir haben außerdem Tipps dazu zusammengestellt, wie sich GenAI je nach Reifegrad und Art der Nutzung auf sichere Weise einsetzen lässt. Damit kann Ihr Unternehmen eine zu seinen Bedürfnissen passende Strategie für die Verwendung von GenAI entwickeln, im Rahmen derer Ihre Daten geschützt sind und die Rechtskonformität gewährleistet ist.

– Dawn Parzych, Director of Product Marketing bei Cloudflare

Geschütztes Experimentieren mit generativer KI



Leider endet Sarahs Geschichte an dieser Stelle. Doch aufgrund der zunehmenden Verbreitung von prädiktiver KI und GenAI wird es in Zukunft unzählige „Sarahs“ geben, die in IT- und Entwicklerteams, als Business-Technologinnen oder an anderer Stelle Einsatz zeigen.

KI fasziniert Fachleute und Alltagsnutzer gleichermaßen und weckt Neugier und Experimentierfreude. Um das volle Potenzial von KI zu erschließen, kommt man um Experimente jedoch nicht herum. Wenn man dabei aber nicht Vorsicht walten lässt und Schutzmaßnahmen trifft, kann es passieren, dass Sicherheitslücken geschaffen werden oder man gegen Vorschriften verstößt.

Um dabei die richtige Balance zu finden und KI-Projekte nicht nur besser zu verstehen, sondern auch effektiver zu steuern, müssen drei Schlüsselbereiche berücksichtigt werden:

1 Einsatz von KI

KI-Technologien (z. B. ChatGPT, Bard und GitHub Copilot) von Drittanbietern nutzen und gleichzeitig Assets (sensible Daten, geistiges Eigentum, Quellcode usw.) schützen und potenzielle Risiken entsprechend des Anwendungsfalls so gering wie möglich halten

2 Entwicklung von KI

Maßgeschneiderte KI-Lösungen entwickeln, die auf die speziellen Bedürfnisse des Unternehmens zugeschnitten sind (z. B. firmeneigenen Algorithmen für vorausschauende Analysen, kundenorientierte Kopiloten oder Chatbots und KI-gestützte Systeme zur Bedrohungserkennung)

3 Absicherung der KI-Revolution

KI-Anwendungen und KI-Systeme vor Manipulationen schützen, die darauf abzielen, dass sie sich auf unvorhergesehene Weise verhalten

Geschütztes Experimentieren mit generativer KI

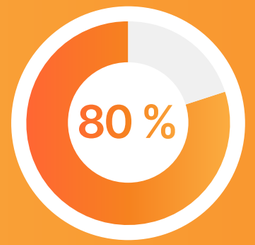
GenAI-Transformation: heute und in Zukunft

GenAI hat bei Verbrauchern und Unternehmen inzwischen eine beispiellose Akzeptanz erreicht. Anfangs gab es nur eine kleine Gruppe von Intentivnutzern, doch die Zahl der Anwender ist schnell gewachsen. Zum Teil war das einer sehr aktiven Open-Source-Community und dem verbrauchergetriebenen Experimentieren mit Apps wie ChatGPT und Stable Diffusion zu verdanken.

Dabei haben die Nutzer festgestellt, dass Roboter den Menschen nicht ersetzen werden.

GenAI versetzt uns vielmehr in die Lage, Dinge zu verfeinern und zu optimieren, anstatt alles von Grund auf neu zu erschaffen. Dadurch kann sie Unternehmen dabei helfen, ihren Mitarbeitenden zu größerer Effizienz zu verhelfen. Vorausschauende KI bietet ähnliche Vorteile, weil sich mit ihrer Hilfe unter anderem leichter Daten zur Verbesserung der Entscheidungsfindung einsetzen, smartere Produkte entwickeln und Kundenerfahrungen personalisieren lassen.

Aktuell nutzen **59 % der Entwickler** KI in ihren Arbeitsabläufen[1]



2026 werden mehr als **80 % der Unternehmen** GenAI-fähige API, Modelle und/oder Anwendungen in Produktivumgebungen einsetzen (aktuell sind es 5 %)[2]

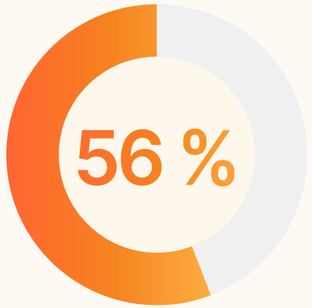
2030 wird GenAI bei **50 % der Aufgaben von Wissensarbeitern** zur Steigerung der Produktivität oder zur Verbesserung der durchschnittlichen Arbeitsqualität genutzt werden (heute ist das bei weniger als 1 % der Fall)[3]

1. SlashData, „How developers interact with AI technologies“, Mai 2024
2. Gartner, „A CTO's Guide to the Generative AI Technology Landscape“, September 2023
3. Gartner, „Emerging Tech: The Key Technology Approaches That Define Generative AI“, September 2023

Sichere Verwendung generativer KI

Das Experimentieren mit KI reicht von der Nutzung vorgefertigter KI-Tools und -Dienste bis hin zur völligen Neuerstellung maßgeschneiderter KI-Lösungen. Manche Unternehmen werden vielleicht irgendwann eigene KI-Modelle und -Anwendungen entwickeln, viele werden jedoch auch in Zukunft KI-Tools von Drittanbietern nutzen.

Doch Letzteres bringt neue Risiken mit sich, weil die Unternehmen ihre Sicherheits- und Datenschutzkonfigurationen nur in begrenztem Umfang direkt kontrollieren.



56 % der Mitarbeitenden haben schon KI-gestützte Tools zur Erledigung ihrer Arbeit verwendet. Doch nur 26 % aller Unternehmen verfügen über eine KI-Richtlinie.[4]

Wahrscheinlich nutzen ihre Mitarbeitenden bereits KI-Tools „von der Stange“. Das können SaaS-Produktreihen wie Microsoft 365 sein oder Chatbots, die in Suchmaschinen oder allgemein zugänglichen Anwendungen integriert sind – oder sogar API.

Um ihrer Sorgfaltspflicht nachzukommen und sich vor den Risiken zu schützen, müssen Unternehmen unter anderem folgende Maßnahmen ergreifen:

- **Bewertung** des Sicherheitsrisikos von Drittanbieter-Tools
- **Zerstreuung** von Datenschutzbedenken
- **Steuerung** der Nutzung (oder übermäßigen Abhängigkeit) von externen API
- **Überwachung** potenzieller Schwachstellen

Ein Beispiel hierfür wäre die Nutzung öffentlicher Webanwendungen wie ChatGPT durch Mitarbeitende. Jeder Prompt wird zu einem Datenelement, das nicht mehr der Kontrolle des Unternehmens unterliegt. Nutzer können sensible, vertrauliche oder gesetzlich regulierte Informationen – wie personenbezogene Daten, Finanzinformationen, geistiges Eigentum oder Quellcode – damit an Unbefugte weitergeben. Und selbst, wenn sensible Informationen nicht direkt offengelegt werden, lassen sich aus den Eingaben eventuell Rückschlüsse auf den Kontext und damit auf solche Daten ziehen.

Zwar existiert eine Einstellung, mit der Mitarbeitende verhindern können, dass das Modell weiter mit ihren Eingaben trainiert wird. Diese muss aber händisch aktiviert werden. Es liegt an den Unternehmen, ihre Mitarbeitenden an der Eingabe vertraulicher Daten zu hindern.

Einschätzung der Gefahren

Datenoffenlegung

Inwieweit geben Nutzer sensible Daten in unzulässiger Weise an externe KI-Dienste weiter? Reichen die genutzten Anonymisierungs-/Pseudonymisierungstechniken aus?

API-Risiken

Wie gehen Sie mit Sicherheitslücken bei Drittanbieter-API um, die Einfallstore für Angreifer sein könnten?

Blackbox-Systeme

Welche Entscheidungsprozesse bei externen KI-Modellen könnten unerwartete Risiken mit sich bringen?

Drittanbieter-Risikomanagement

Was wissen Sie über die Sicherheitsmaßnahmen Ihrer KI-Anbieter? Wichtiger noch: Was wissen Sie nicht?

4. The Conference Board, September 2023

Schutzmaßnahmen bei der Verwendung von KI

1 Datenkontrolle steuern und Risiken eindämmen

- Entwickeln von Richtlinien für die Art und den Zeitpunkt der KI-Nutzung (Informationen, deren Weitergabe mit GenAI das Unternehmen den Nutzern gestattet, Richtlinien für die Zugriffskontrolle, Compliance-Anforderungen, Meldung von Verstößen usw.)
- Folgenabschätzung zur Sammlung von Informationen und zur Ermittlung und Quantifizierung der Vorteile und Risiken der KI-Nutzung

2 Überblick verbessern und Kontrolle über Sicherheit und Datenschutz erhöhen

- Protokollieren aller Verbindungen (auch zu KI-Anwendungen) zur kontinuierlichen Überwachung der Nutzeraktivitäten, der Verwendung von KI-Tools und der Datenzugriffsmuster sowie zur Erkennung von Anomalien
- Ermitteln der vorhandenen Schatten-IT (einschließlich KI-Tools) und über das Genehmigen, Blockieren oder Einführen zusätzlicher Kontrollen entscheiden
- Durchsuchen der Konfigurationen von SaaS-Anwendungen auf Sicherheitsrisiken (z. B. OAuth-Berechtigungen, die nicht autorisierten KI-fähigen Anwendungen von zugelassenen Applikationen erteilt werden, wodurch die Offenlegung von Daten droht)

3 Bei KI-Tools ein- und ausgehende Daten kontrollieren, um das herauszufiltern, was geistiges Eigentum oder die Vertraulichkeit von Daten gefährden oder gegen Urheberrecht verstoßen könnte

- Anwenden von Sicherheitskontrollen zur Festlegung der Art und Weise, in der Nutzer mit KI-Tools interagieren können (z. B. Uploads stoppen, Kopieren/Einfügen verhindern und nach sensiblen/geschützten Daten suchen und deren Eingabe blockieren)
- Einführen von Sicherheitsvorkehrungen, die KI-Bots daran hindern, Ihre Website zu durchsuchen
- Vollständiges Blockieren von KI-Tools nur, wenn keine anderen Kontrollen möglich sind: Nutzer finden sonst immer Umgehungsmöglichkeiten, die sich Ihrer Kontrolle entziehen

4 Zugriffskontrolle für KI-Anwendungen und -Infrastruktur

- Strenge Identitätsprüfung für jeden Nutzer und jedes Gerät beim Zugriff auf KI-Tools zur Festlegung der zulässigen Anwender
- Implementieren identitätsbasierter Zero Trust-Zugriffskontrollen unter Anwendung des Prinzips der minimalen Rechtevergabe zur Schadensbegrenzung im Falle von Kontenkompromittierungen oder Insider-Bedrohungen

5 Kosten und Betriebseffizienz optimieren

- Mithilfe von Analysen und Protokollen mehr über die Verwendung von KI-Applikationen in Erfahrung bringen, um bei stärkerer Nutzung die Kontrolle über Durchsatzbegrenzung, Zwischenspeicherung und erneute Anfragen zu haben und ein Alternativ-Modell erstellen zu können

Schutz der Produktentwicklungen

Trainieren Ihres KI-Modells

Durch KI-Pipelines entstehen neue Schwachstellen. Aufgrund unserer Erfahrung mit der Absicherung zu Beginn und während des gesamten Entwicklungsprozesses haben wir aber eine gute Vorstellung davon, wie man sich vor diesen schützen kann. Es liegt nahe, bei der Gewährleistung der KI-Sicherheit bei Ihrem Modell anzusetzen.

Grundsätzlich kann alles, was zum Training eines KI-Modells genutzt wird, im Output der KI-Anwendung auftauchen. Deshalb sollten Sie sich zunächst überlegen, wie Sie diese Daten absichern können, um später keine unangenehmen Überraschungen zu erleben. Wenn Sie Ihre Daten nicht schützen, besteht die Gefahr, dass sich Ihre Angriffsfläche vergrößert und das später bei der Anwendung Probleme verursacht.

Um die Möglichkeiten für eine vorsätzliche oder versehentliche Datenkompromittierung auf ein Mindestmaß zu begrenzen, muss der Schutz der Datenintegrität unbedingt gewährleistet sein. Zu den Sicherheitsrisiken in der KI-Pipeline gehören:

- **Datenvergiftung:** Schädliche Datensätze beeinträchtigen das Endergebnis und verursachen Verzerrungen
- **Missbrauch von Halluzinationen:** Feindliche Akteure lassen KI-Halluzinationen – bei denen die Maschine Informationen erfindet, um eine Antwort liefern zu können – als legitim erscheinen, sodass schädliche und unzulässige Datensätze die Ergebnisse verfälschen können

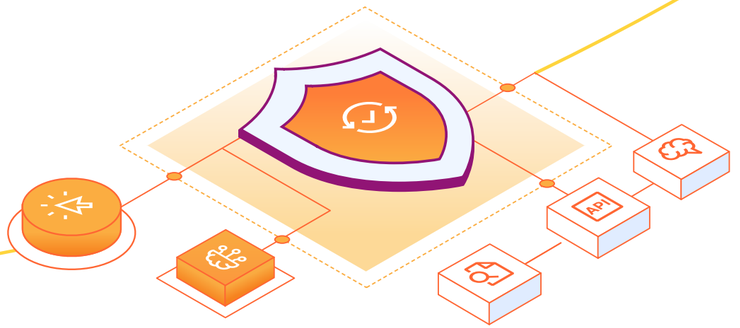
Wenn Sie keine eigenen Modelle trainieren, besteht der erste Schritt in der Auswahl eines Modells zur Erledigung von Aufgaben. In diesem Fall sollten Sie sich anschauen, wie das Modell entwickelt wurde und geschützt ist, weil das für die Inferenz eine Rolle spielt.

Inferenz ist der Prozess, der auf das Training der KI folgt. Je besser ein Modell trainiert und je feiner es abgestimmt ist, desto besser funktioniert die Inferenz – auch wenn sie nie perfekt ist. Selbst bestens trainierte Modelle können halluzinieren.

Sicherheit nach der Implementierung

Wenn Sie Ihre interne KI-Anwendung erstellt und implementiert haben, müssen Sie ihre vertraulichen Daten schützen und den Zugriff darauf absichern. Neben den bereits genannten Empfehlungen (wie dem Durchsetzen der Token-Verwendung für jeden Nutzer und eine Durchsatzbegrenzung), sollten Sie Folgendes in Betracht ziehen:

- **Kontingentverwaltung:** Grenzwerte verhindern die Kompromittierung und Weitergabe von API-Schlüsseln der Nutzer
- **Blockierung bestimmter Autonomous System Numbers (ASN):** Verhindert das Senden zu großer Mengen an Datenverkehr durch Angreifer
- **Warteräume oder Nutzer-Tests:** Macht Anfragen schwieriger oder zeitaufwendiger, sodass sich diese für Angreifer nicht mehr rentieren
- **Erstellung und Validierung eines API-Schemas:** Beschreibt die beabsichtigte Nutzung durch Identifizierung und Katalogisierung aller API-Endpunkte und führt dann alle spezifischen Parameter und Typgrenzen auf
- **Analyse der Tiefe und Komplexität von Abfragen:** Hilft beim Schutz vor direkten DoS-Angriffen und Entwicklerfehlern, gewährleistet einen guten Zustand Ihres Ursprungsservers und eine erwartungsgemäße Zustellung von Anfragen an Ihre Nutzer
- **Konsequenter Zugriff per Sicherheits-Token:** Schützt vor Zugriffskompromittierung, wenn Token auf Middleware-Ebene oder im API-Gateway validiert werden



Robuster Bedrohungsschutz für Experimente mit generativer KI

Beim Experimentieren mit GenAI sollte in jeder Phase – von der ersten Einführung bis zur vollständigen Implementierung – nur ein minimales oder tolerierbares Risiko bestehen. Dieser Leitfaden bietet Ihnen das nötige Rüstzeug, um Ihre digitale Umgebung zu kontrollieren – unabhängig davon, ob Ihr Unternehmen KI in irgendeiner Form nutzt, entwickelt oder dies plant.

Mit der Einführung neuer Funktionen zu zögern, ist normal. Doch es gibt Ressourcen, die es Ihnen erlauben, in aller Gelassenheit und auf sichere Weise mit KI zu experimentieren. Davon benötigen Unternehmen heute am dringendsten eine Möglichkeit, IT und Sicherheit zu vernetzen. Dieses Bindeglied dient als eine Art roter Faden und reduziert die Komplexität, weil es mit allem in der Umgebung zusammenarbeitet, überall verfügbar ist und die erforderlichen Sicherheits-, Netzwerks- und Entwicklungsfunktionen ausführt.

Dadurch können Sie sich folgender Dinge gewiss sein, wenn Sie GenAI nutzen:

- Einhaltung von Vorschriften und die Fähigkeit, die Übermittlung gesetzlich regulierter Daten zu erkennen und zu kontrollieren
- Wiedererlangung der Übersicht und Kontrolle über sensible Daten in SaaS-Anwendungen, Schatten-IT und neuen KI-Tools
- Schutz von Entwicklercode durch das Erkennen und Blockieren von Quellcode beim Hoch- und Herunterladen; Verhindern, Aufspüren und Beheben von Fehlkonfigurationen in SaaS-Anwendungen und Cloud-Diensten, einschließlich Code-Repositorys

Da sich KI ständig weiterentwickelt, ist Unsicherheit vorprogrammiert. Aus diesem Grund ist ein stabilisierender Einfluss wie der von Cloudflare von großem Vorteil.

Schutz vor KI-Risiken durch drei Arten von LLM

Je nach Nutzung variiert das mit KI einhergehende Risiko. Daher müssen sich Unternehmen über die verschiedenen mit der Nutzung und Entwicklung von Large Language Models (LLM) verbundenen Gefahren informieren, um sich anschließend bei der Implementierung von LLM aktiv einbringen zu können.

LLM-Kategorie	Hauptrisiko
Intern	Zugang zu sensiblen Daten und geistigem Eigentum
Produkt	Imageschaden
Öffentlich	Verlust sensibler Daten

Skalierbarkeit, Benutzerfreundlichkeit und nahtlose Integration

Die Connectivity Cloud von Cloudflare gibt Ihnen die Kontrolle und ermöglicht eine bessere Übersicht und größeren Schutz, damit Sie auf sichere Weise mit KI experimentieren und gegebenenfalls skalieren können. Dank unserer Dienste müssen sie keine Abwägung zwischen Nutzererfahrung und Sicherheit treffen.

Die meisten Unternehmen sind entweder reine Anwender von KI, oder sie nutzen sie und entwickeln gleichzeitig ein eigenes Produkt. In jedem Fall bedeutet der Einsatz von Cloudflare, dass ihre KI-Projekte nie ins Stocken geraten.

- Mit unserem **globalen Netzwerk** sind Kontrollmaßnahmen an jedem erforderlichen Ort schnell skalier- und durchsetzbar
- Unsere **Benutzerfreundlichkeit** erleichtert die Implementierung und Verwaltung von Richtlinien zur KI-Nutzung
- Dank einer einzigen, **programmierbaren Architektur** können Applikationen durch mehrschichtige Sicherheitsvorkehrungen ohne Störung der KI-Verwendung geschützt werden

Die Connectivity Cloud von Cloudflare schützt alle Aspekte Ihrer KI-Experimente, insbesondere:

- Unsere **Zero Trust- und Secure Access Service Edge (SASE)**-Dienste tragen zur Reduzierung der mit der **Nutzung** von KI-Tools von Drittanbietern verbundenen Risiken auf ein Mindestmaß bei
- Unsere **Entwicklerplattform** hilft Ihrem Unternehmen bei der sicheren und effizienten **Erstellung** eigener KI-Werkzeuge und -Modelle
- Im Rahmen einer **KI-gestützten Absicherung** trägt unsere Plattform mit KI-Verfahren und maschinellem Lernen Bedrohungsinformationen zusammen, die zum Schutz von Unternehmen bei KI-Experimenten genutzt werden

	Wie Sie KI nutzen	Wie Sie KI erstellen
Die Größe unseres globalen Netzwerks	Kontrollen werden überall einheitlich skaliert und durchgesetzt	Inferenz, Abfragen und Zwischenspeicherung werden beschleunigt
Unsere einfache Verwaltung	Eine einzige Steuerungsebene mit einfacher Implementierung und Richtlinien	Vorlagen für eine schnelle Einführung
Unsere einheitliche und programmierbare Netzwerkarchitektur	Neue Sicherheitsmaßnahmen können ohne Störung der KI-Nutzung eingeführt werden	Integrierter Datenschutz und integrierte Compliance

Nächste Schritte

Ob Absicherung des Einsatzes von KI in Ihrem Unternehmen oder Schutz der von Ihnen entwickelten KI-Anwendungen: Cloudflare for AI erfüllt Ihre Bedürfnisse. Mit unseren Diensten können Sie neue Funktionen in beliebiger Reihenfolge mit grenzenloser Interoperabilität und flexibler Integration einführen.

Mit einem Experten sprechen

Weitere Informationen finden Sie hier:
cloudflare.com

+49 89 2555 2276 | enterprise@cloudflare.com | www.cloudflare.com/de-de/

REV:BDES-6307.2024AUG19