

WHITE PAPER

# Freitextsuche in AXIS Camera Station Pro

Januar 2025

# Zusammenfassung

AXIS Camera Station Pro enthält mehrere forensische Video-Suchwerkzeuge. Dazu gehören Zeitleisten-Scrubbing, Datensuche und Intelligente Suche mit vorklassifizierten Objekten und Freitextsuche.

Mit der Freitextsuche können Sie nach beliebigen beweglichen Objekten suchen, indem Sie sie mit Ihren eigenen Worten beschreiben. Durch die Möglichkeit, detaillierte Suchfilter mit einer Vielzahl von beschreibenden Attributen zu erstellen, können Sie relevantes Filmmaterial schneller auffinden.

Die Freitextsuchfunktion basiert auf einem Text-Bild-Abgleich, der von einem vorab trainierten Open-Source-Foundation Model bereitgestellt wird, das von Axis für Videosicherheitsanwendungen optimiert wurde. Die Suche kann auf eine Kamera oder mehrere Kameras gleichzeitig angewendet werden.

Eine numerische Darstellung Ihrer Freitextabfrage wird mit numerischen Darstellungen von Bildern erfasster bewegter Objekte verglichen. Die Ergebnisse werden als Miniaturbilder angezeigt, einschließlich Kameraname, Uhrzeit und Datum, sortiert nach Relevanz für Ihre Suchanfrage.

Bei der Freitextsuche setzen wir KI ein, um die Genauigkeit und Effizienz unserer Lösungen für die forensische Suche zu erhöhen und letztlich die menschliche Entscheidungsfindung zu verbessern. Zur Einhaltung rechtlicher und ethischer Standards enthält die Suchfunktion eine separate, von Axis entwickelte Moderationsfunktion, die die Verwendung anstößiger Wörter in Suchanfragen einschränkt. Alle Suchanfragen werden protokolliert und Administratoren können sie einsehen, sodass Missbrauch nachvollzogen und entsprechend korrigierend eingegriffen werden kann.

# Inhalt

1	Einführung	4
2	Hintergrund: Forensische Suche in AXIS Camera Station Pro	4
3	Wie funktioniert die Freitextsuche?	4
4	Aufbau von Suchanfragen	5
	4.1 Richtlinien für Prompts	6
	4.2 Moderation von Abfragen	6
	4.3 Protokollierung der Suchanfragen	6
5	Text-Bild-Abgleich	6
	5.1 Für die Videosicherheit optimiertes Foundation Model	7
	5.2 Feature Vectors	7
6	Prozessablauf	7
7	Verantwortungsvoller Einsatz der KI	9

# 1 Einführung

Die Freitextsuche in AXIS Camera Station Pro ermöglicht es Ihnen, Videoaufzeichnungen mit Ihren eigenen Worten statt mit vordefinierten Filtern zu durchsuchen.

In diesem Whitepaper wird die Funktionsweise der Suchmethode erläutert und es werden einige Leitlinien zu ihrer Anwendung vorgestellt. Wir beschreiben auch die Moderationsfunktion und die Abfrageprotokollierung, die die Einhaltung rechtlicher und ethischer Standards gewährleisten.

## 2 Hintergrund: Forensische Suche in AXIS Camera Station Pro

AXIS Camera Station Pro ist mit mehreren vorinstallierten forensischen Video-Suchwerkzeugen ausgestattet, wie Zeitleisten-Scrubbing, Datensuche sowie eine intelligente Suche mit vorklassifizierten Objekten und Freitextsuche.

Die intelligente Suche verwendet Szenenmetadaten, die vom Axis Gerät generiert wurden. Zu den Metadaten gehören der Typ des Objekts (Person, Fahrzeugtyp oder unbekanntes Objekt) bei beweglichen Objekten sowie Merkmale wie Kleidungs- und Fahrzeugfarbe, Lizenzkennzeichen, Geschwindigkeit, Position und Zeitstempel.

Bei Geräten mit eingeschränkten Analysefunktionen basiert die Suchfunktion auf der Bewegungserkennung im Gerät in Kombination mit der Objektklassifizierung, die auf dem AXIS Camera Station Pro Server ausgeführt wird. Die forensische Suche in AXIS Camera Station Pro ist somit eine hybride Lösung, bei der nach Möglichkeit vorrangig die Edge-Geräte genutzt und bei Bedarf durch Daten vom Server ergänzt werden.

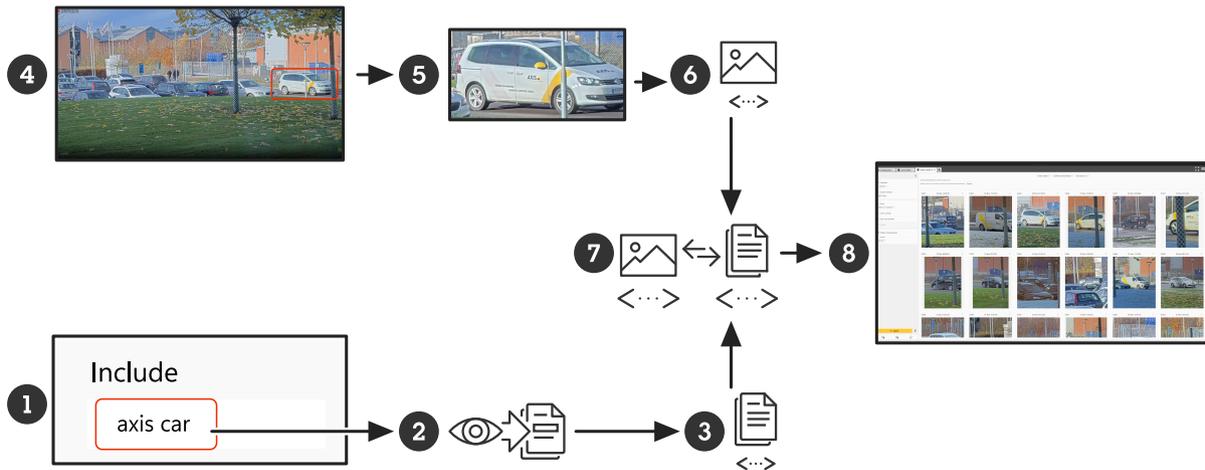
Bisher musste die Suche nach Szenenmetadaten mit vordefinierten Suchfiltern durchgeführt werden. Bei diesen wählt man feste Objektbeschreibungen aus einer Liste, einschließlich Objekttyp (z. B. „Fahrzeug“), Fahrzeugtyp (sofern zutreffend, z. B. „Auto“), Farbe (z. B. „blau“) und mehr. Bei der neuen Freitextsuche können Sie stattdessen Ihren eigenen Suchfilter erstellen.

Die vorklassifizierte Suche liefert zwar hochpräzise Ergebnisse, kann aber neuartige Typen von Objekten, die nicht vordefiniert sind, nicht erkennen. Die Freitextsuche bietet Ihnen die Freiheit und Flexibilität, mit Ihren eigenen Worten zu suchen. Sie können jedes bewegliche Objekt mit natürlicher Sprache und Assoziationen detaillierter beschreiben, um Ihre Suche zu verfeinern und mehr Ergebnisse zu erhalten.

## 3 Wie funktioniert die Freitextsuche?

Eine numerische Darstellung Ihrer Freitextabfrage wird mit numerischen Darstellungen von Bildern erfasster bewegter Objekte verglichen. Das Ergebnis dieses Text-Bild-Vergleichs wird angezeigt und nach

der besten Übereinstimmung sortiert. Die Ergebnisse werden als Miniaturbilder angezeigt, einschließlich Kameraname, Uhrzeit und Datum, sortiert nach Relevanz für Ihre Suchanfrage.



Vereinfachte Übersicht der Freitextsuche. Bitte beachten Sie, dass die Schritte 4 bis 6 kontinuierlich ablaufen, auch wenn gerade keine Suche stattfindet, um Feature Vectors aller erkannten bewegten Objekte zu erstellen.

- 1 Sie tippen Ihre Suchanfrage als Freitext ein.
- 2 Ein Moderationsmodul verhindert die Verwendung von toxischen und unethischen Wörtern.
- 3 Das Foundation Model erstellt eine numerische Darstellung (einen Feature Vector) der Suchanfrage.
- 4 Eine Kamera erkennt Bewegungsabläufe in einer Szene.
- 5 Die Kamera wählt einen Bildausschnitt aus, um das bewegte Objekt darzustellen.
- 6 Das Foundation Model erstellt einen Feature Vector des Objekts, nachdem es dessen Form, Muster, Farbe usw. analysiert hat.
- 7 Die beiden Feature Vectors werden verglichen.
- 8 Das Ergebnis des Vergleichs wird nach der besten Übereinstimmung sortiert und als Miniaturansicht dargestellt.

Die Freitextsuche kann auf eine Kamera oder mehrere Kameras gleichzeitig angewendet werden.

Um den Umfang Ihrer Freitextsuche einzugrenzen, können Sie sie mit anderen intelligenten Suchfunktionen wie der Ähnlichkeitssuche oder der zeitbasierten Suche kombinieren, indem Sie nacheinander mehrere Suchtypen ausprobieren.

## 4 Aufbau von Suchanfragen

Sie können nach jedem beweglichen Objekt und jedem Typ von Fahrzeug suchen. Befolgen Sie die Richtlinien, um optimale Ergebnisse zu erzielen.

Bitte beachten Sie, dass Sie nur nach bewegten Objekten suchen sollten. Die Suche nach stationären Objekten wird in den meisten Fällen erfolglos sein.

Suchphrasen werden moderiert und protokolliert, um unethisches Suchverhalten zu verhindern.

## 4.1 Richtlinien für Prompts

- Beschreiben Sie Situationen so, wie Sie ein Bild beschreiben würden. Das Modell wird mit Standbildern gefüttert, weshalb die Suche nach Aktionen (wie Hinfallen, Laufen oder Stehlen) schwierig sein kann, da diese mehr Kontext erfordern.
- Beschreiben Sie Objekte mit einigen wenigen Schlüsseldeskriptoren: „eine Person in einem roten Pullover“ oder „ein gelber Pickup“. Wie andere multimodale Modelle funktioniert das Freitextsuchmodell gut mit Deskriptoren wie Objekten und Farben, ist aber weniger geeignet für Personenzählungen („drei Personen“), Dialekte oder emotionale Hinweise („wütend aussehender Mann“). Die Objektbeschreibung sollte weder subjektiv noch zu vage sein oder zu viele spezifische Details enthalten.
- Kombinieren Sie mehrere Objektattribute mit *und*: „Person mit rotem Hut und Rucksack“.
- Beschreiben Sie Text, Textlogos oder Markennamen: „Lieferwagen mit Text Axis“.
- Konzentrieren Sie sich nicht auf die Beschreibung von Umgebungen. Die Verarbeitung erfolgt an zugeschnittenen Bildern von Objekten, sodass das Modell die Umgebung der Objekte möglicherweise gar nicht zu sehen bekommt. Allgemeine Szenen- oder Umgebungsbeschreibungen (wie „Stadt“, „urban“, „Park“, „Garten“, „See“ oder „Strand“) führen daher möglicherweise nicht zu guten Ergebnissen.
- Experimentieren Sie mit alternativen Formulierungen, wenn Sie mit einem Ergebnis nicht zufrieden sind.
- Die Freitextabfrage unterstützt nur Englisch.

## 4.2 Moderation von Abfragen

Wir haben eine Abfragemoderation auf der Grundlage gängiger Praktiken eingeführt, um eine effektive Filterung zu gewährleisten. Das Moderationsmodell ist ein Modell zur Verarbeitung natürlicher Sprache, das die Abfrage überprüft, um anstößige Formulierungen einzuschränken. Es überprüft Textinhalte auf schädliche, unangemessene oder beleidigende Inhalte. Außerdem haben wir diese Funktionen um eigene Maßnahmen erweitert, einschließlich einer benutzerdefinierten Liste von verbotenen Suchkategorien und -wörtern. Wenn eine Suchanfrage bestimmte Wörter oder Ausdrücke enthält, blockieren wir die Suche, um eine sichere und verantwortungsvolle Erfahrung zu gewährleisten. Wenn Sie mit der Blockierung eines Wortes nicht einverstanden sind oder ein Wort blockiert haben möchten, können Sie uns anonym Feedback geben.

## 4.3 Protokollierung der Suchanfragen

AXIS Camera Station Pro erstellt einen Audit-Trail der Aktionen des Benutzers. Audit-Trails protokollieren nicht nur die einzelnen Aktivitäten und die Identität des Benutzers, sondern auch alle im Betrieb verwendeten Daten. Es werden sämtliche Suchvorgänge des Benutzers protokolliert, einschließlich der Suchanfragen. Anhand der Protokolle können Administratoren unangemessenes Suchverhalten von Benutzern erkennen, ethisch bedenkliche Suchanfragen markieren und entsprechende Korrekturmaßnahmen einleiten.

Die Videodaten werden nicht mit Axis geteilt. Ihre Daten bleiben auf Ihrem Server.

# 5 Text-Bild-Abgleich

Die Möglichkeit, Videometadaten mit Hilfe von Freitextabfragen zu durchsuchen, erweitert die Suchmöglichkeiten von einer vordefinierten Liste von Attributen auf nahezu unbegrenzte Suchkriterien. In

AXIS Camera Station Pro basiert diese Funktion auf einem Open-Source-Foundation Model, das anhand von Milliarden von Bild-Text-Paaren trainiert und von Axis für den Einsatz in der Videosicherheit optimiert wurde, um deren Leistung zu verbessern.

## 5.1 Für die Videosicherheit optimiertes Foundation Model

Das Foundation Model ist ein Text-Bild-Modell, das auf großen Datensätzen von Text-Bild-Paaren trainiert wurde. Es handelt sich um ein Zero-Shot-Modell, das Text mit relevanten Bildern abgleicht. Ein Zero-Shot-Modell ist ein Typ von Modell in der künstlichen Intelligenz (KI), das Objekte oder Konzepte ohne vorherige Trainingsdaten erkennen und klassifizieren kann. Mit anderen Worten: Das Modell kann Aufgaben ausführen, ohne zuvor Beispiele für diese Aufgabe gesehen zu haben. Diese Fähigkeit ist entscheidend, um eine optimale Leistung beim Abgleich von natürlicher Sprache mit Bildern zu gewährleisten.

Das Modell wurde an einer großen Menge von Text-Bild-Kombinationen trainiert und arbeitet mit einem neuronalen Netzwerk mit mehr als 2,5 Milliarden Parametern. Bei Axis haben wir dieses Modell mit eigenem Schulungsmaterial weiter optimiert, um Bilder mit typischen Ansichten von Sicherheitskameras und Objekten besser interpretieren zu können. Wir haben das Modell also speziell für Videosicherheitsanwendungen optimiert.

## 5.2 Feature Vectors

Wenn Sie eine Freitextsuche durchführen, erstellt das Foundation Model (Basismodell) einen Feature Vector der Suchanfrage.

Das Foundation Model erstellt außerdem fortlaufend Beschreibungen aller Objekte, die in den Szenen getrackt werden, und erstellt Feature Vectors zu deren Darstellung. Jedes Objekt wird durch nur einen Feature Vector repräsentiert, der in unserer Datenbank gespeichert wird. Dies ermöglicht eine schnelle Suche, da die Feature Vectors bereits vorberechnet und in der Datenbank leicht verfügbar sind.

Beide Typen von Feature Vectors werden in die Vektorvergleichsmaschine eingespeist, um die Ähnlichkeit/Distanz zwischen Ihrer Suchanfrage und allen verfügbaren Feature Vectors der erfassten Objekten zu bestimmen.

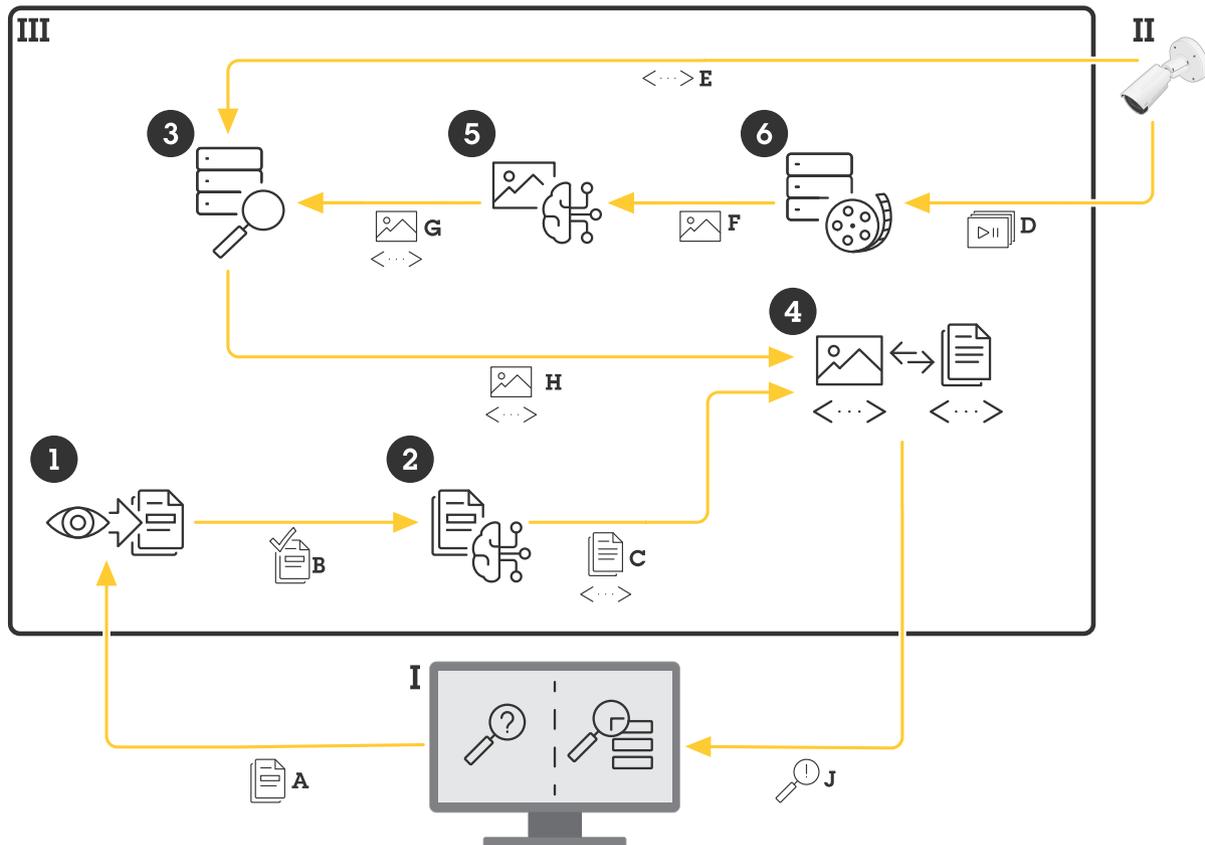
Ein Feature Vector ist eine numerische Darstellung von Text oder Bildern. Die Feature Vectors von Personen oder Objekten sind also nur abstrakte Repräsentationen des Aussehens der Person oder des Objekts. Feature Vectors enthalten keine vom Menschen interpretierbaren Informationen über Merkmale wie z. B. Haar- oder Kleidungsfarbe, die explizit einer bestimmten Person zugeordnet oder zur Identifizierung verwendet werden können. Die Feature Vectors können nur für Vergleiche mit anderen Feature Vectors verwendet werden.

# 6 Prozessablauf

Das Prozessdiagramm gibt einen Überblick über die wichtigsten Schritte, inklusive Standorten und den generierten Datentypen.

Zu beachten ist, dass die obere Schleife im Diagramm, bestehend aus der Kamera (II), dem Speicher für Aufzeichnungen (6), dem Foundation Model (5) und der Suchdatenbank (3), ein kontinuierlich ablaufender

Prozess ist, der Feature Vectors aller detektierten bewegten Objekte erstellt – und zwar nicht nur, wenn Sie eine Suche durchführen.



#### Haupteinsatzorte (I – III) für die Freitextsuche

- I AXIS Camera Station Pro Client
- II Kamera(s)
- III AXIS Camera Station Pro Server

#### Wichtigste Prozessschritte (1 – 6)

- 1 Moderation von Suchanfragen
- 2 Foundation Model (Text)
- 3 Suche in der Datenbank
- 4 Vektorvergleich
- 5 Foundation Model (Bild)
- 6 Speicher für Aufzeichnungen

#### Datentyp oder Ergebnis (A – J)

- A Zeichenkette
- B Zeichenkette
- C Feature Vector (Text)
- D Video
- E Metadaten
- F Bilder
- G Feature Vectors (Bild)
- H Feature Vectors (Bild)
- J Ergebnisse der Suche

(I) **AXIS Camera Station Pro Client:** Hier geben Sie Ihre Suchanfrage ein und erhalten sortierte Suchergebnisse

(II) **Kamera(s):** Die Freitextsuche funktioniert bei Axis Kameras mit AXIS OS 5.51 oder höher, doch je besser die Kamera ist, desto bessere Ergebnisse erhalten Sie. Ältere Geräte liefern weniger detaillierte Metadaten, die nur auf Bewegungserkennung basieren. Neuere Geräte erzeugen AXIS Scene Metadaten mit einer Objektklassifizierung. Die Erfassung und das Tracking von beweglichen Objekten durch die Kamera wird genutzt, um ein repräsentatives Bild jedes erfassten Objekts zu finden und so die Anzahl der auf dem Server zu analysierenden Bilder zu reduzieren.

(III) **AXIS Camera Station Pro Server:** Hier werden alle Metadaten und Videodaten der Kameras verarbeitet und gespeichert. Bevor Sie eine Freitextsuche durchführen, muss der Server (für jedes erfasste bewegte Objekt) das Video dekodieren und ein Bild des erfassten Objekts extrahieren. Daraufhin verarbeitet das Foundation Model dieses Bild und erstellt den Feature Vector. Diese Operationen benötigen eine ziemlich große Menge an Rechenleistung, so dass zur Verbesserung der Leistung die Feature Vectors in einer Datenbank gespeichert werden, was eine schnelle zukünftige Suche ermöglicht. Wenn Ihr Server über freie Kapazitäten verfügt, empfehlen wir dringend, die Hintergrundverarbeitung der Videos Ihrer wichtigsten Kameras zu aktivieren, da dies die Suche erheblich beschleunigt.

(1) **Moderation von Suchanfragen:** Das Moderationsmodell überprüft die Abfrage auf anstößige Inhalte.

(2) **Foundation Model (Text):** Das Foundation Model erstellt eine numerische Repräsentation (Feature Vector) des moderierten Suchanfragen-Textstrings.

(3) **Suche in der Datenbank:** Die Suchdatenbank enthält vollständige Metadaten aus AXIS Scene Metadata oder vom Server erstellte Metadaten, einschließlich Objektklassifizierungsdaten mit Attributen, Uhrzeit, Position und Feature Vectors.

(4) **Vektorieller Vergleich:** Die Feature Vector-Darstellung der Textzeichenfolge aus der Suchanfrage wird mit den Feature Vector-Darstellungen der im Video erfassten Objekte verglichen.

(5) **Foundation Model (Bild):** Das Foundation Model erstellt numerische Darstellungen (Feature Vectors) der einzelnen Objekt-Tracks im aufgezeichneten Video. Dies ist ein kontinuierlicher Prozess, der auch dann abläuft, wenn Sie nichts suchen.

(6) **Speicher für Aufzeichnungen:** Hier werden die Videos der Kamera gespeichert, von hier erhält das Foundation Model seine Bilder.

## 7 Verantwortungsvoller Einsatz der KI

Bei der Freitextsuche setzen wir KI ein, um die Genauigkeit und Effizienz unserer Lösungen für die forensische Suche zu erhöhen und letztlich die menschliche Entscheidungsfindung zu verbessern.

Verantwortlichkeit und Rechenschaftspflicht sind von grundlegender Bedeutung für den Ansatz von Axis im Bereich der künstlichen Intelligenz. Dabei müssen wir sicherstellen, dass die von uns entwickelten KI-Systeme ethischen Grundsätzen entsprechen, die Gesetze einhalten und Risiken effektiv managen. Axis stellt Werkzeuge zur Verfügung, mit denen unsere Kunden die Integrität ihrer Betriebsabläufe sicherstellen können. Die Freitextsuche in AXIS Camera Station Pro umfasst ein fein abgestimmtes Textklassifizierungsmodell für die Moderation von Textprompts. Wir haben dieses Modell entwickelt, um Suchanfragen zu moderieren, damit Sie eine unethische Nutzung leichter verhindern können.

Die Freitextsuche verbindet sich einmal wöchentlich mit den Axis Cloud-Diensten, um zu überprüfen, ob die KI-Modelle aktualisiert werden müssen, um neuen Vorschriften oder Anforderungen gerecht zu werden. Wenn die Verbindung fehlschlägt, ist keine Freitextsuche möglich, bis die Verbindung wiederhergestellt ist.

Um die Einhaltung rechtlicher und ethischer Standards bei der Anwendung von KI zu gewährleisten, bieten unsere Produkte Zugriffskontrollen auf Basis von Zugangsdaten zur Authentifizierung von Benutzern und Zugriffsberechtigungen. Dies ermöglicht unseren Kunden, die Einhaltung der Betriebsrichtlinien durch die Benutzer durchzusetzen.



# Über Axis Communications

Axis erhöht die Sicherheit und Geschäftsperformance – mit intelligenten Lösungen für eine sichere Welt. Als Branchenführer in Sachen Netzwerktechnologie bietet Axis Lösungen in den Bereichen Videosicherheit und Zutrittskontrolle sowie Türsprechanlagen und Audiosysteme. Die branchenweit anerkannten Schulungen der Axis Communications Academy vermitteln fundiertes Expertenwissen zu den neuesten Technologien.

Axis beschäftigt rund 4000 Personen in über 50 Ländern und arbeitet eng mit Technologiepartnern und Systemintegratoren zusammen, um stets die beste Lösung anzubieten. Axis wurde 1984 gegründet und hat seinen Hauptsitz in Lund, Schweden.