

Swobodne wyszukiwanie tekstowe w oprogramowaniu **AXIS Camera Station Pro**

Styczeń 2025

Streszczenie

AXIS Camera Station Pro oferuje kilka fabrycznie instalowanych narzędzi do przeszukiwania materiału wizyjnego na potrzeby prac wyjaśniających. Obejmują one scrubbing osi czasu, przeszukiwanie danych oraz Inteligentne wyszukiwanie przy użyciu gotowych klasyfikacji obiektów i swobodnego wyszukiwania tekstowego.

Swobodne wyszukiwanie tekstowe umożliwia wyszukiwanie dowolnych poruszających się obiektów przez opisanie ich własnymi słowami. Duża dowolność tworzenia szczegółowych filtrów wyszukiwania przy użyciu szerokiej gamy atrybutów opisowych umożliwia szybsze znajdowanie odpowiednich części nagranych materiału wizyjnego.

Funkcja swobodnego wyszukiwania tekstowego opiera się na dopasowywaniu tekstu i obrazu za pomocą fabrycznie przeszkolonego modelu open source, który firma Axis zoptymalizowała pod kątem dozoru. Wyszukiwanie można zastosować do jednej kamery lub kilku kamer naraz.

Numeryczna reprezentacja zapytania z tekstem swobodnym jest porównywana z numeryczną reprezentacją obrazów wygenerowanych dzięki detekcji poruszających się obiektów. Wyniki są wyświetlane jako miniatury z nazwą kamery, godziną i datą posortowane według dokładności dopasowania do zapytania wyszukiwania.

Do obsługi funkcji swobodnego wyszukiwania tekstowego wykorzystywana jest sztuczna inteligencja (AI), co ma na celu zwiększenie dokładności i wydajności naszych rozwiązań do wyszukiwania na potrzeby prac wyjaśniających, a w ostatecznym rozrachunku również usprawnienie podejmowania decyzji przez człowieka. W celu zachowania zgodności ze standardami prawnymi i etycznymi ta funkcja wyszukiwania obejmuje oddzielną, opracowaną przez Axis funkcję moderacji, która ogranicza używanie obraźliwych słów w zapytaniach wyszukiwania. Ponadto wszystkie wyszukiwania są rejestrowane i widoczne dla administratorów, co umożliwia podejmowanie dalszych kroków i przeprowadzanie działań naprawczych w przypadku ewentualnego niewłaściwego korzystania z tego narzędzia.

Spis treści

1	Wprowadzenie	4
2	Zagadnienie ogólne: wyszukiwanie w oprogramowaniu AXIS Camera Station Pro na potrzeby prac wyjaśniających	4
3	Jak działa swobodne wyszukiwanie tekstowe?	4
4	Tworzenie zapytań wyszukiwania	5
	4.1 Wskazówki dotyczące zapytań	6
	4.2 Moderacja zapytań	6
	4.3 Rejestrowanie zapytań wyszukiwania	6
5	Dopasowywanie tekstu do obrazu	6
	5.1 Model fundamentowy zoptymalizowany pod kątem systemów dozoru	7
	5.2 Wektory cech	7
6	Przegląd procesu	7
7	Odpowiedzialne korzystanie ze sztucznej inteligencji	9

1 Wprowadzenie

Narzędzie do swobodnego wyszukiwania tekstowego w ramach AXIS Camera Station Pro umożliwia przeszukiwanie nagrań wizyjnych przy użyciu własnych słów, zamiast fabrycznie zdefiniowanych filtrów.

W tym dokumencie opisano sposób działania tej metody wyszukiwania i kilka wskazówek dotyczących korzystania z niej. Przedstawiono tu też funkcje moderacji oraz rejestrowania zapytań, które mają na celu zapewnianie zgodności ze standardami prawnymi i etycznymi.

2 Zagadnienie ogólne: wyszukiwanie w oprogramowaniu AXIS Camera Station Pro na potrzeby prac wyjaśniających

Oprogramowanie AXIS Camera Station Pro obejmuje kilka fabrycznie instalowanych narzędzi do wyszukiwania materiału wizyjnego na potrzeby prac wyjaśniających, w tym do scrubbingu osi czasu, przeszukiwania danych, a także Inteligentnego wyszukiwania przy użyciu uprzednio sklasyfikowanych obiektów i swobodnego wyszukiwania tekstowego.

Inteligentne wyszukiwanie wykorzystuje metadane sceny generowane przez urządzenie Axis. Metadane te obejmują typ obiektu (osoba, typ pojazdu lub obiekt nieznanymi) w przypadku poruszających się obiektów, wraz z ich atrybutami, takimi jak ubranie, kolor pojazdu, numery rejestracyjne, prędkość, lokalizacja i znacznik czasowy.

W przypadku urządzeń z ograniczonymi możliwościami analitycznymi funkcja wyszukiwania bazuje na detekcji ruchu w urządzeniu oraz klasyfikacji obiektów dokonywanej na serwerze AXIS Camera Station Pro. Wyszukiwanie na potrzeby prac wyjaśniających w oprogramowaniu AXIS Camera Station Pro jest zatem rozwiązaniem hybrydowym, w którym jak najlepiej wykorzystuje się możliwości urządzeń brzegowych, a w razie potrzeby uzupełnia się je o dane z serwera.

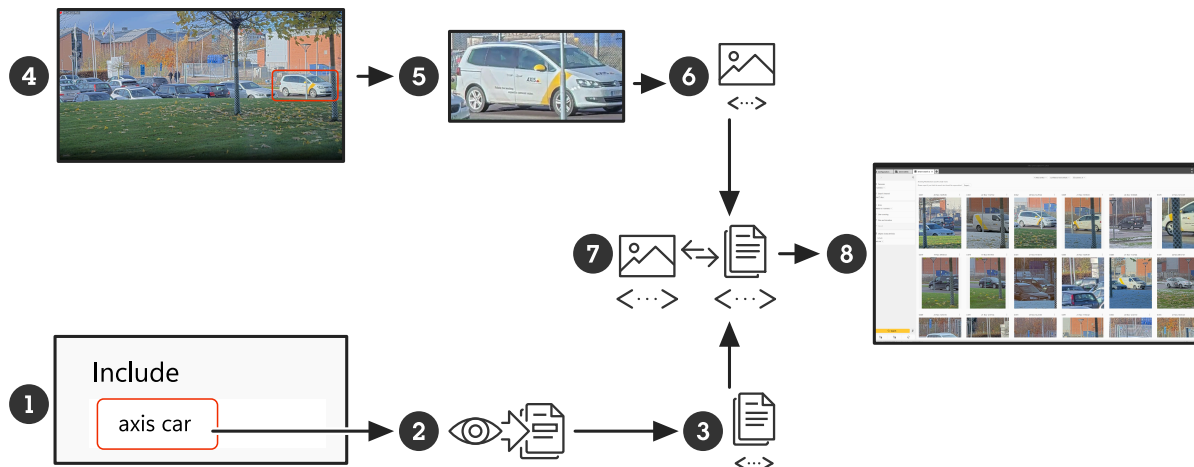
Tradycyjnie wyszukiwanie przy użyciu metadanych sceny musiało się odbywać przy użyciu fabrycznie zdefiniowanych filtrów wyszukiwania. Za ich pomocą można było wybrać gotowe deskryptory obiektów z listy, w tym typ obiektu (np. „pojazd”), typ pojazdu (jeśli dotyczy, np. „samochód”), kolor (np. „niebieski”) itd. Tymczasem nowa metoda swobodnego wyszukiwania tekstowego umożliwia tworzenie własnych filtrów wyszukiwania.

Wyszukiwanie przy użyciu gotowej klasyfikacji wprawdzie zapewnia wysoką precyzję wyników, jednak nie pozwala na wykrywanie nowych typów obiektów, które nie zostały fabrycznie zdefiniowane. Aby wyeliminować to ograniczenie, swobodne wyszukiwanie tekstowe zapewnia dużą dowolność i elastyczność wyszukiwania przy użyciu własnych słów. Każdy poruszający się obiekt można opisać bardziej szczegółowo za pomocą języka naturalnego i skojarzeń, aby doprecyzować wyszukiwanie i uzyskać więcej wyników.

3 Jak działa swobodne wyszukiwanie tekstowe?

Numeryczna reprezentacja zapytania z tekstem swobodnym jest porównywana z numeryczną reprezentacją obrazów wygenerowanych dzięki detekcji poruszających się obiektów. Wyniki dopasowania tekstu do obrazu są prezentowane i sortowane w kolejności od najlepszego dopasowania. Wyniki są wyświetlane jako

miniatury z nazwą kamery, godziną i datą posortowane według dokładności dopasowania do zapytania wyszukiwania.



Uproszczony przegląd procesu swobodnego wyszukiwania tekstowego. Należy pamiętać, że kroki 4–6 są wykonywane w sposób ciągły (nawet gdy nie przeprowadzasz wyszukiwania) w celu tworzenia wektorów cech wszystkich wykrytych poruszających się obiektów.

- 1 Użytkownik wpisuje zapytanie wyszukiwania z tekstem swobodnym.
- 2 Moduł moderacji zapobiega używaniu szkodliwych i nieetycznych słów.
- 3 Model fundamentowy tworzy numeryczną reprezentację (wektor cech) zapytania wyszukiwania.
- 4 Kamera wykrywa ruch w danej scenie.
- 5 Kamera wybiera jeden przycięty obraz reprezentujący poruszający się obiekt.
- 6 Model fundamentowy tworzy wektor cech obiektu po przeanalizowaniu jego kształtu, wzorów, koloru i innych atrybutów.
- 7 Przeprowadzane jest porównanie tych dwóch wektorów cech.
- 8 Wyniki porównania są sortowane w kolejności od najlepszego dopasowania i prezentowane w postaci miniatur.

Swobodne wyszukiwanie tekstowe można zastosować do jednej kamery lub kilku kamer naraz.

Aby zawęzić zakres swobodnego wyszukiwania tekstowego, można je połączyć z innymi funkcjami Inteligentnego wyszukiwania, takimi jak wyszukiwanie podobieństw czy wyszukiwanie na podstawie czasu – wystarczy użyć jeden typ wyszukiwania po drugim.

4 Tworzenie zapytań wyszukiwania

Możesz wyszukać dowolny poruszający się obiekt i dowolnego typu pojazd. Aby uzyskać najlepsze wyniki, postępuj zgodnie z niniejszymi wskazówkami.

Pamiętaj, że wyszukiwanie powinno dotyczyć tylko poruszających się obiektów. Wyszukiwanie nieruchomych obiektów na ogół nie zadziała.

Szukane sformułowania są moderowane i rejestrowane w celu zapobiegania nieetycznym zachowaniom.

4.1 Wskazówki dotyczące zapytań

- Opisz sytuacje tak, jak opisuje się obraz. Model jest zasilony zdjęciami, więc wyszukiwanie czynności (takich jak upadek, bieg czy kradzież) może być trudne, ponieważ wymagałoby więcej kontekstu.
- Opisz obiekty za pomocą kilku kluczowych deskryptorów: „osoba w czerwonym swetrze” lub „żółta furgonetka”. Podobnie jak inne modele multimodalne, model swobodnego wyszukiwania tekstowego dobrze sobie radzi z takimi deskryptorami jak obiekty i kolory, ale mniej się sprawdza w przypadku liczb („trzy osoby”), slangu czy określeń emocji („rozzłoszczony mężczyzna”). Opis obiektu nie powinien być subiektywny, zbyt niejasny ani zawierać zbyt szczegółowych informacji.
- Łącz wiele atrybutów obiektów za pomocą operatora *i*: „osoba nosząca czerwoną czapkę i plecak”.
- Uwzględniaj w opisie napisy, logo tekstowe lub nazwę marki: „van z napisem Axis”.
- Nie skupiaj się na opisywaniu otoczenia. Przetwarzanie odbywa się na podstawie przyciętych obrazów obiektów, co oznacza, że model może nie mieć danych na temat ich otoczenia. Szersze znaczeniowo deskryptory sceny lub otoczenia (takie jak „miasto”, „miejski”, „park”, „ogród”, „jezioro” czy „plaża”) mogą zatem nie przekładać się na dobre wyniki.
- Jeśli wyniki Cię nie zadowolają, próbuj używać innych sformułowań.
- Zapytania z tekstem swobodnym są obsługiwane tylko w języku angielskim.

4.2 Moderacja zapytań

Wdrożyliśmy moderację zapytań zgodnie z ogólnie przyjętą praktyką mającą na celu zapewnienie skutecznego filtrowania. Model moderacji jest modelem przetwarzania języka naturalnego, który sprawdza zapytanie w celu ograniczenia obraźliwych sformułowań. Sprawdza on całe ciągi tekstowe pod kątem krzywdzących, niestosownych lub szkodliwych treści. Ponadto rozszerzyliśmy te funkcje o własne rozwiązania, w tym własną listę zabronionych podczas wyszukiwania kategorii i słów. Jeśli zapytanie zawiera słowa lub sformułowania z tej listy, zostaje odrzucone w celu zachowania bezpieczeństwa środowiska wyszukiwania i zadbania o etyczność jego wyników. Jeśli nie zgadzasz się z zablokowaniem lub chcesz zasugerować zablokowanie jakiegoś słowa, możesz wysłać Axis anonimową opinię.

4.3 Rejestrowanie zapytań wyszukiwania

AXIS Camera Station Pro zachowuje ścieżkę audytu operacji użytkownika. Ścieżki audytu zawierają nie tylko konkretne operacje i tożsamość użytkownika, ale i wszelkie dane użyte w ramach tych operacji. Oznacza to, że są rejestrowane wszystkie operacje wyszukiwania użytkownika, w tym zapytania wyszukiwania. Administratorzy mogą używać tych rejestrów do identyfikowania niewłaściwych zachowań użytkowników podczas wyszukiwania, wykrywania nieetycznych zapytań wyszukiwania i podejmowania działań naprawczych.

Należy zauważyć, że dane wizyjne nie są udostępniane Axis. Dane pozostają na serwerze klienta.

5 Dopasowywanie tekstu do obrazu

Możliwość przeszukiwania metadanych materiału wizyjnego za pomocą zapytań z tekstem swobodnym znacznie zwiększa możliwości wyszukiwania, umożliwiając przejście od predefiniowanej listy atrybutów do niemal nieograniczonych kryteriów wyszukiwania. W oprogramowaniu AXIS Camera Station Pro funkcja ta jest oparta na modelu fundamentowym open source, przeszkolonym przy użyciu miliardów

par obraz-tekst i starannie dopracowanym przez Axis pod kątem użytkowania w systemach dozoru w celu poprawy efektywności.

5.1 Model fundamentowy zoptymalizowany pod kątem systemów dozoru

Model fundamentowy jest modelem dopasowywania tekstu do obrazu, który został przeszkolony przy użyciu dużych zbiorów danych par tekst-obraz. Jest to model zero-shot, który dopasowuje tekst do odpowiednich obrazów. Model zero-shot to rodzaj modelu sztucznej inteligencji (AI), który jest w stanie rozpoznawać i klasyfikować obiekty lub pojęcia bez wcześniejszych danych szkoleniowych. Innymi słowy, model ten może wykonywać zadania bez uprzednich przykładów związanych z danym zadaniem. Zdolność ta ma kluczowe znaczenie z perspektywy zapewniania optymalnej wydajności i skuteczności dopasowywania języka naturalnego do obrazów.

Model został przeszkolony przy użyciu dużej ilości kombinacji tekst-obraz i operuje na sieci neuronowej ponad 2,5 miliarda parametrów. W Axis wykorzystaliśmy nasz własny, unikatowy materiał szkoleniowy w celu dalszego dopracowania tego modelu, zwiększając jego zdolność do interpretowania obrazów z typowymi obszarami obserwacji kamer do nadzoru i obiektami. Oznacza to, że zoptymalizowaliśmy model pod kątem systemów dozoru.

5.2 Wektory cech

W ramach swobodnego wyszukiwania tekstowego model fundamentowy tworzy wektor cech zapytania wyszukiwania.

Oprócz tego model fundamentowy stale generuje opisy każdego obiektu śledzonego w scenach i tworzy wektory cech służące do ich reprezentowania. Każdy obiekt jest reprezentowany za pomocą tylko jednego wektora cech, który jest przechowywany w naszej bazie danych. Przekłada się na to na dużą szybkość wyszukiwania, ponieważ potrzebne wektory cech są już obliczone i dostępne w bazie danych.

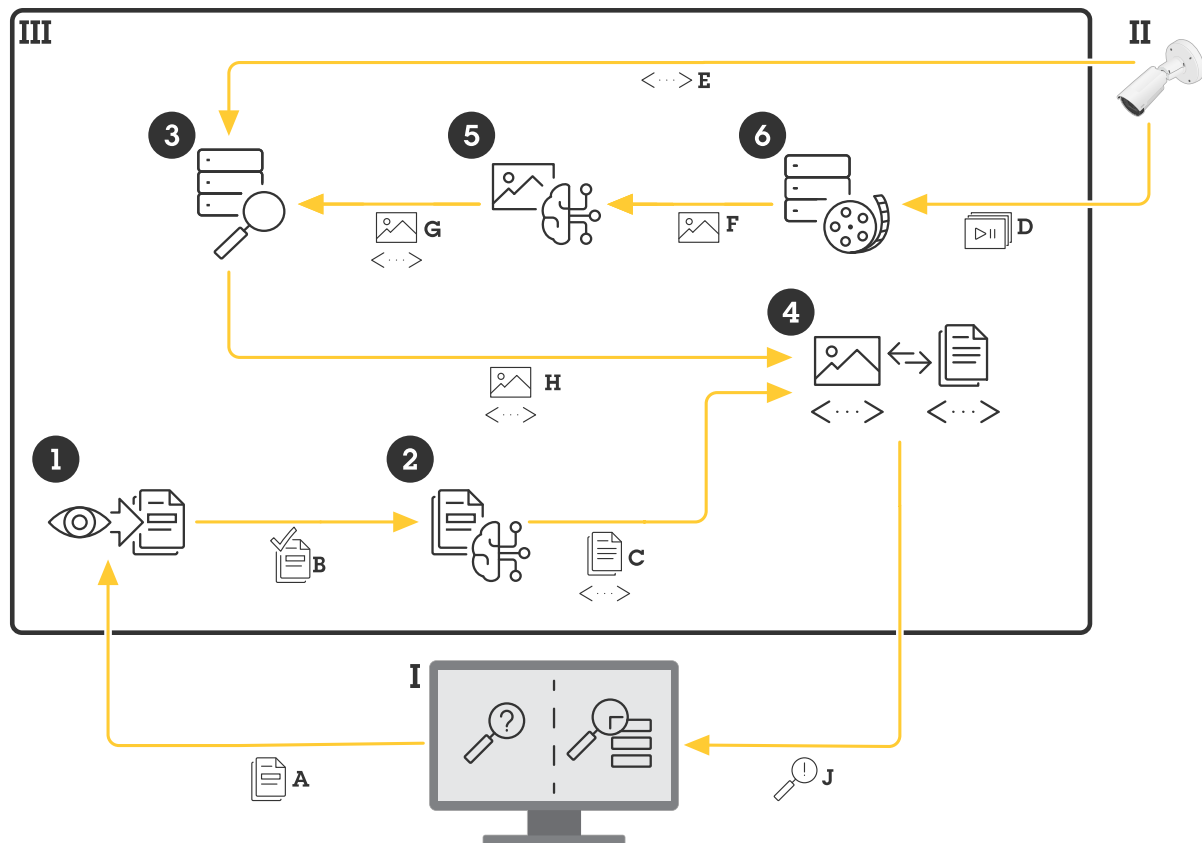
Oba typy wektorów cech są wprowadzane do silnika porównywania wektorów w celu ustalenia odległości podobieństwa między zapytaniem wyszukiwania a wszystkimi dostępnymi wektorami cech, które reprezentują wykryte obiekty.

Wektor cech jest numeryczną reprezentacją tekstu lub obrazu. Wektor cech osoby lub obiektu jest zatem jedynie abstrakcyjną reprezentacją wyglądu osoby lub obiektu. Wektory cech nie zawierają żadnych nadających się interpretacji przez człowieka informacji o cechach, takich jak kolor włosów czy elementu ubrania, które można by jednoznacznie przypisać do konkretnej osoby bądź wykorzystać do identyfikacji. Wektory cech służą tylko do porównań z innymi wektorami cech.

6 Przegląd procesu

Schemat przeglądu procesu przedstawia główne etapy procesu, w tym miejsca realizacji etapów i rodzaj danych generowanych przez każdy z nich.

Należy zauważyć, że pętla u góry schematu, obejmująca kamerę (II), pamięć masową z nagraniami (6), model fundamentowy (5) i bazę danych wyszukiwania (3), jest procesem wykonywanym stale – w celu tworzenia wektorów cech wszystkich wykrytych poruszających się obiektów – a nie tylko podczas wyszukiwania.



Główne miejsca realizacji (I-III) procesu swobodnego wyszukiwania tekstowego

- I Klient AXIS Camera Station Pro
- II Kamery
- III Serwer AXIS Camera Station Pro

Główne etapy procesu (1-6)

- 1 Moderacja zapytań wyszukiwania
- 2 Model fundamentowy (tekst)
- 3 Baza danych wyszukiwania
- 4 Porównanie wektorów
- 5 Model fundamentowy (obraz)
- 6 Pamięć masowa z nagraniami

Typ danych lub wynik (A-J)

- A Ciąg tekstowy
- B Ciąg tekstowy
- C Wektor cech (tekst)
- D Nagranie wideo
- E Metadane
- F Obrazy
- G Wektory cech (obraz)
- H Wektory cech (obraz)
- J Wyniki wyszukiwania

(I) Klient AXIS Camera Station Pro: tutaj wpisuje się zapytanie i otrzymuje posortowane wyniki wyszukiwania.

(II) Kamery: swobodne wyszukiwanie tekstowe działa w przypadku kamer Axis z systemem AXIS OS 5.51 lub nowszym, ale im lepsza kamera, tym lepsze wyniki można uzyskać. Starsze urządzenia zapewniają mniej szczegółowe metadane wyłącznie na podstawie detekcji ruchu. Nowsze urządzenia generują dane aplikacji AXIS Scene Metadata, które obejmują klasyfikację obiektów. Detekcja i śledzenie przez kamerę poruszających się obiektów są wykorzystywane do znalezienia jednego reprezentatywnego obrazu każdego wykrytego obiektu, co przekłada się na mniej obrazów do przeanalizowania na serwerze.

(III) Serwer AXIS Camera Station Pro: tutaj są przetwarzane i przechowywane wszystkie metadane i dane wizyjne z kamer. Zanim będzie można przeprowadzić swobodne wyszukiwanie tekstowe, serwer musi (w przypadku każdego wykrytego poruszającego się obiektu) zdekodować materiał wizyjny i wyodrębnić obraz wykrytego obiektu. Następnie model fundamentowy przetwarza ten obraz w celu utworzenia wektora cech. Operacje te wymagają dość dużej mocy obliczeniowej, więc w celu poprawienia wydajności wektory cech są zapisywane w bazie danych, co umożliwi późniejsze szybkie wyszukiwanie. Jeśli serwer ma nieco wolnej mocy, zdecydowanie zalecamy włączenie przetwarzania materiału wizyjnego z najważniejszych kamer w tle, ponieważ to znacznie przyspiesza wyszukiwanie.

(1) Moderacja zapytań wyszukiwania: model moderacji sprawdza zapytanie w celu ograniczenia obraźliwych treści.

(2) Model fundamentowy (tekst): model fundamentowy tworzy numeryczną reprezentację (wektor cech) moderowanego ciągu tekstowego zapytania wyszukiwania.

(3) Baza danych wyszukiwania: baza danych wyszukiwania zawiera pełne metadane z aplikacji AXIS Scene Metadata lub metadane utworzone przez serwer, w tym dane dotyczące klasyfikacji obiektów przy użyciu atrybutów, czasu, położenia i wektorów cech.

(4) Porównanie wektorów: reprezentacja wektora cech ciągu tekstowego zapytania wyszukiwania jest porównywana z reprezentacjami wektorów cech obiektów wykrytych w materiale wizyjnym.

(5) Model fundamentowy (obraz): model fundamentowy tworzy numeryczne reprezentacje (wektory cech) każdej ścieżki obiektu w nagrany materiał wizyjnym. Jest to proces ciągły, wykonywany również wtedy, gdy nie jest przeprowadzana operacja wyszukiwania.

(6) Pamięć masowa z nagraniami: jest to miejsce przechowywania materiału wizyjnego z kamer i źródło obrazów dla modelu fundamentowego.

7 Odpowiedzialne korzystanie ze sztucznej inteligencji

Do obsługi funkcji swobodnego wyszukiwania tekstowego wykorzystywana jest sztuczna inteligencja (AI), co ma na celu zwiększenie dokładności i wydajności naszych rozwiązań do wyszukiwania na potrzeby prac wyjaśniających, a w ostatecznym rozrachunku również usprawnienie podejmowania decyzji przez człowieka.

Odpowiedzialność stanowi fundament podejścia Axis do AI. Wiąże się ona z dbaniem o to, by tworzone przez nas systemy AI były zgodne z zasadami etycznymi i przepisami prawa oraz skutecznie zarządzały ryzykiem. Axis udostępnia narzędzia, dzięki którym klienci mogą mieć pewność integralności swoich operacji. Funkcja swobodnego wyszukiwania tekstowego w oprogramowaniu AXIS Camera Station Pro obejmuje precyzyjnie dopasowany model klasyfikacji tekstu w celu moderowania zapytań tekstowych. Model ten opracowaliśmy z myślą o moderowaniu zapytań wyszukiwania w celu zapobiegania nieetycznym zastosowaniom.

Funkcja swobodnego wyszukiwania tekstowego raz w tygodniu łączy się z usługami Axis w chmurze, aby sprawdzić, czy modele AI wymagają uaktualnienia w celu zapewnienia zgodności z nowo wprowadzonymi przepisami lub wymogami. Jeśli połączenie się nie powiedzie, nie będzie można korzystać ze swobodnego wyszukiwania tekstowego do czasu ponownego nawiązania połączenia przez system.

Aby zapewniać zgodność ze standardami prawnymi i etycznymi dotyczącymi zastosowań sztucznej inteligencji, nasze produkty oferują kontrolę dostępu na podstawie danych uwierzytelniających użytkownika i uprawnień dostępu. Dzięki temu nasi klienci mogą egzekwować przestrzeganie przez użytkowników zasad działania.

O firmie Axis Communications

Axis wspiera rozwój inteligentnego oraz bezpiecznego świata przez tworzenie rozwiązań umożliwiających poprawę bezpieczeństwa i efektywności biznesowej. Jako firma zajmująca się technologiami sieciowymi oraz lider branży Axis oferuje rozwiązania z zakresu dozoru wizyjnego, kontroli dostępu, interkomów i systemów audio. Ich rozszerzeniem i uzupełnieniem są inteligentne aplikacje analityczne oraz wysokiej jakości szkolenia.

Axis zatrudnia około 4000 pracowników w ponad 50 krajach oraz współpracuje z partnerami z obszaru technologii i integracji systemów na całym świecie w celu dostarczania swoich rozwiązań klientom. Firma została założona w 1984 roku i ma swoją siedzibę w Lund w Szwecji.