

LIVRE BLANC

# Recherche en texte libre dans **AXIS Camera Station Pro**

Janvier 2025

## Avant-propos

AXIS Camera Station Pro inclut plusieurs outils préinstallés de recherche vidéo forensique : parmi eux, le « scrubbing » de la chronologie, la recherche de données et la recherche intelligente, avec objets préclassés et recherche en texte libre.

La recherche en texte libre vous permet de rechercher n'importe quel objet mobile en le décrivant avec vos propres mots. La possibilité de créer des filtres de recherche détaillés avec un large éventail d'attributs descriptifs permet de trouver plus rapidement des séquences vidéo pertinentes.

La fonction de recherche en texte libre est basée sur la correspondance texte-image fournie par un modèle de fondation open source préalablement entraîné et optimisé par Axis pour les scénarios de surveillance. La recherche peut s'appliquer à une ou plusieurs caméras en même temps.

Une représentation numérique de votre requête en texte libre est comparée aux représentations numériques des images d'objets mobiles détectés. Les résultats sont affichés sous forme de vignettes, avec le nom de la caméra, l'heure et la date, triés par pertinence par rapport à votre requête de recherche.

Avec la recherche en texte libre, nous utilisons l'IA pour augmenter la précision et l'efficacité de nos solutions de recherche forensique et, au final, améliorer la prise de décision humaine. Afin de respecter les normes juridiques et éthiques, la fonction de recherche comprend une fonction de modération distincte, développée par Axis, qui limite l'usage de mots offensants dans les requêtes de recherche. Toutes les recherches sont également consignées dans un journal et consultables par les administrateurs, ce qui permet d'assurer un suivi et de prendre des mesures correctives en cas d'usage abusif.

# Table des matières

1	Introduction	4
2	Contexte : recherche forensique dans AXIS Camera Station Pro	4
3	Principe de fonctionnement de la recherche en texte libre	4
4	Élaboration d'une requête de recherche	5
	4.1 Conseils pour les invites de recherche	6
	4.2 Modération des requêtes	6
	4.3 Journalisation des requêtes de recherche	6
5	Correspondance texte-image	6
	5.1 Modèle de fondation optimisé pour la surveillance	7
	5.2 Vecteurs de caractéristiques	7
6	Aperçu du processus	7
7	Usage responsable de l'IA	9

# 1 Introduction

L'outil de recherche en texte libre d'AXIS Camera Station Pro vous permet de rechercher des enregistrements vidéo avec vos propres mots au lieu de filtres prédéfinis.

Ce livre blanc décrit le fonctionnement de la méthode de recherche et présente quelques directives sur ses modalités d'emploi. Nous décrivons également la fonction de modération et la journalisation des requêtes en place pour garantir le respect des normes juridiques et éthiques.

## 2 Contexte : recherche forensique dans AXIS Camera Station Pro

AXIS Camera Station Pro inclut plusieurs outils de recherche vidéo forensique préinstallés, notamment scrubbing de la chronologie, recherche de données, mais aussi recherche intelligente avec des objets préclassés et la recherche en texte libre.

La fonction de recherche intelligente utilise les métadonnées de scène générées par le dispositif Axis. Les métadonnées comprennent le type d'objet (personne, type de véhicule ou objet inconnu) pour les objets en mouvement, ainsi que des attributs tels que la couleur des vêtements et des véhicules, les plaques d'immatriculation, la vitesse, la position et un horodatage.

Pour les dispositifs dotés de capacités d'analyse limitées, la fonction de recherche est basée sur la détection des mouvements dans le dispositif, combinée à la classification des objets effectuée sur le serveur AXIS Camera Station Pro. La recherche forensique dans AXIS Camera Station Pro est donc une solution hybride où les capacités des dispositifs sont exploitées dans toute la mesure du possible, mais complétées le cas échéant par des données provenant du serveur.

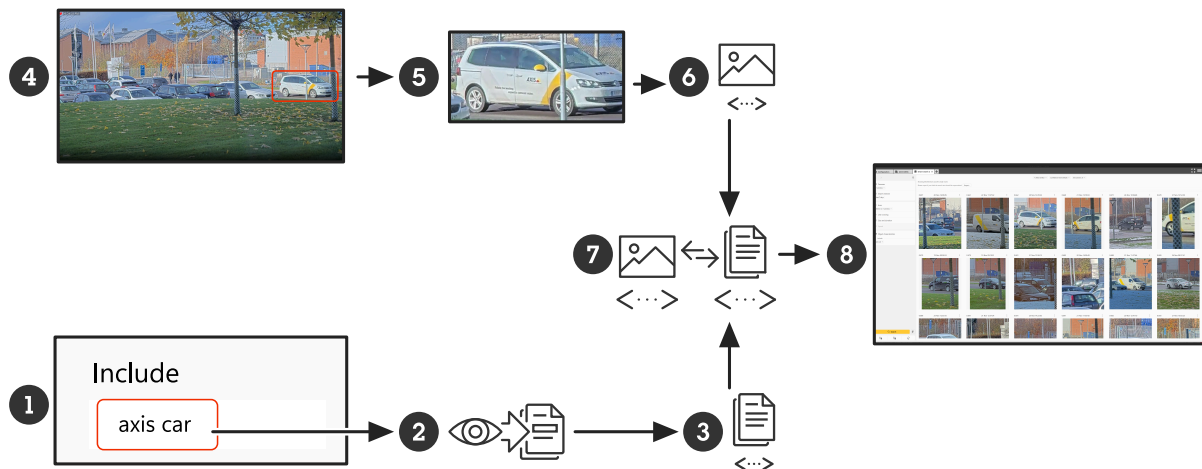
Habituellement, les recherches utilisant les métadonnées de scène devaient être effectuées à l'aide de filtres de recherche prédéfinis. Avec ces derniers, vous choisissez des descripteurs d'objet fixes dans une liste, parmi lesquels le type d'objet (tel que « véhicule »), le type de véhicule (le cas échéant, par exemple « voiture »), la couleur (telle que « bleu ») et plus encore. À l'inverse, la nouvelle méthode de recherche en texte libre vous permet de créer votre propre filtre de recherche.

Si la recherche par pré-classification donne des résultats très précis, elle ne peut pas détecter les nouveaux types d'objet qui ne sont pas prédéfinis. Pour remédier à cette limitation, la recherche en texte libre vous offre la liberté et la flexibilité de rechercher en utilisant vos propres mots. Vous pouvez décrire tout objet en mouvement de manière plus détaillée en langage naturel et avec des associations pour affiner votre recherche et obtenir davantage de résultats.

## 3 Principe de fonctionnement de la recherche en texte libre

Une représentation numérique de votre requête en texte libre est comparée aux représentations numériques des images d'objets mobiles détectés. Le résultat de cette comparaison texte-image est présenté et trié par

pourcentage de correspondance. Les résultats sont affichés sous forme de vignettes, avec le nom de la caméra, l'heure et la date, triés par pertinence par rapport à votre requête de recherche.



Aperçu simplifié du processus de recherche en texte libre. Notez que les étapes 4 à 6 se déroulent en continu, même lorsque vous n'effectuez pas de recherche, afin de créer des vecteurs de caractéristiques de tous les objets mobiles détectés.

- 1 Vous tapez votre requête de recherche en texte libre.
- 2 Un module de modération empêche l'utilisation d'expressions répréhensibles et contraires à l'éthique.
- 3 Le modèle de fondation crée une représentation numérique (un vecteur de caractéristiques) de la requête de recherche.
- 4 Une caméra détecte du mouvement dans une scène.
- 5 La caméra sélectionne une image recadrée pour représenter l'objet en mouvement.
- 6 Le modèle de fondation crée un vecteur de caractéristiques de l'objet après avoir analysé sa forme, ses motifs, sa couleur, etc.
- 7 Les deux vecteurs de caractéristiques sont comparés.
- 8 Le résultat de la comparaison est trié par pourcentage de correspondance et présenté sous forme de vignettes.

La recherche en texte libre peut s'appliquer à une ou plusieurs caméras en même temps.

Pour réduire le périmètre de votre recherche en texte libre, vous pouvez la combiner à d'autres fonctions de recherche intelligente, telles que la recherche de similitudes ou la recherche par fenêtre de temps, en utilisant successivement plusieurs types de recherche.

## 4 Élaboration d'une requête de recherche

Vous pouvez rechercher tout objet en mouvement et tout type de véhicule. Suivez les instructions pour obtenir les meilleurs résultats.

Notez que vous ne devez rechercher que des objets en mouvement. La recherche d'objets immobiles sera infructueuse dans la plupart des cas.

Les expressions de recherche sont modérées et consignées afin d'éviter les comportements de recherche contraires à l'éthique.

## 4.1 Conseils pour les invites de recherche

- Décrivez les situations comme vous décririez une image. Comme le modèle est alimenté par des images fixes, la recherche d'actions (telles que tomber, courir ou voler) peut s'avérer difficile car elle nécessiterait davantage de contexte.
- Décrivez les objets avec quelques descripteurs clés : « personne portant un pull rouge » ou « pick-up jaune ». Comme d'autres modèles multimodaux, le modèle de recherche en texte libre donne de bons résultats avec des descripteurs tels que des objets et des couleurs, mais il est moins adapté pour le comptage (« trois personnes »), l'argot ou les perceptions émotionnelles (« homme à l'air furieux »). La description de l'objet ne doit pas être subjective, trop vague ou inclure des détails trop spécifiques.
- Combinez plusieurs attributs d'objet à l'aide de et : « personne avec chapeau rouge et sac à dos ».
- Décrivez un texte, un logo ou une marque : « fourgon avec texte Axis ».
- Ne vous focalisez pas sur la description d'environnements. Le traitement étant effectué sur des images d'objet recadrées, le modèle ne connaît pas nécessairement l'environnement des objets. Des descripteurs généraux de la scène ou de l'environnement (tels que « ville », « urbain », « parc », « jardin », « lac » ou « plage ») risquent donc de ne pas donner de bons résultats.
- Essayez d'autres formulations si vous n'êtes pas satisfait du résultat.
- L'invite en texte libre accepte l'anglais uniquement.

## 4.2 Modération des requêtes

Nous avons mis en place un mécanisme de modération des requêtes basé sur les pratiques admises afin de garantir un filtrage efficace. Le modèle de modération est un modèle de traitement du langage naturel qui vérifie la requête afin de limiter les formulations offensantes. Il vérifie que les chaînes de texte entières ne contiennent aucun contenu indigne, inapproprié ou dangereux. En outre, nous avons amélioré ces capacités grâce à des mesures exclusives, notamment une liste personnalisée de catégories de recherche et de mots interdits. Lorsqu'une requête contient des mots ou des expressions figurant sur cette liste, nous rejetons la recherche afin de maintenir un environnement de recherche sûr et de garantir des résultats éthiques. En tant qu'utilisateur, vous pouvez communiquer anonymement vos commentaires à Axis si vous voulez contester ou suggérer le blocage d'un mot.

## 4.3 Journalisation des requêtes de recherche

AXIS Camera Station Pro conserve un journal d'audit des opérations des utilisateurs. Les journaux d'audit permettent non seulement de garder une trace de chaque opération et identité utilisateur associée, mais aussi de conserver toutes les données utilisées dans l'opération. Cela signifie que toutes les recherches de l'utilisateur, y compris les invites de recherche, sont enregistrées. Les administrateurs peuvent utiliser les journaux pour identifier les comportements de recherche inappropriés chez les utilisateurs, repérer les invites de recherche contraires à l'éthique et prendre des mesures correctives.

Notez que les données vidéo ne sont pas partagées avec Axis. Vos données restent sur votre serveur.

# 5 Correspondance texte-image

La recherche de métadonnées vidéo à l'aide de requêtes en texte libre élargit considérablement le champ des possibles, qui passe d'une liste prédéfinie d'attributs à des critères de recherche presque illimités. Dans

AXIS Camera Station Pro, cette fonction repose sur un modèle de fondation open source, entraîné sur des milliards de paires image-texte et affiné par Axis pour les scénarios de surveillance afin d'améliorer les performances.

## 5.1 Modèle de fondation optimisé pour la surveillance

Le modèle de fondation est un modèle texte-image entraîné sur de vastes ensembles de données de paires texte-image. Il s'agit d'un modèle d'apprentissage « Zero-shot » qui fait correspondre du texte à des images pertinentes. Un modèle Zero-shot est un type de modèle d'intelligence artificielle (IA) capable de reconnaître et de classer des objets ou des concepts sans données d'apprentissage préalables. Concrètement, le modèle peut effectuer des tâches sans en avoir vu d'exemples auparavant. Cette capacité est cruciale pour garantir des performances optimales dans les associations entre langage naturel et images.

Le modèle a été entraîné sur un grand nombre de combinaisons texte-image et fonctionne sur un réseau de neurones comptant plus de 2,5 milliards de paramètres. Chez Axis, nous avons utilisé nos propres ressources d'entraînement pour affiner ce modèle et améliorer sa capacité à interpréter des images, avec des vues de caméras et des objets typiques rencontrés dans le domaine de la surveillance. Cela signifie que nous avons optimisé le modèle pour les scénarios de surveillance.

## 5.2 Vecteurs de caractéristiques

Lorsque vous effectuez une recherche en texte libre, le modèle de fondation crée un vecteur de caractéristiques de la requête de recherche.

Le modèle de fondation produit également en permanence des descriptions de chaque objet suivi dans les scènes et crée des vecteurs de caractéristiques pour les représenter. Chaque objet est représenté par un seul vecteur de caractéristiques, qui est stocké dans notre base de données. La recherche est donc rapide puisque les vecteurs de caractéristiques sont déjà précalculés et facilement disponibles dans la base de données.

Ces deux types de vecteurs de caractéristiques sont introduits dans le moteur de comparaison de vecteurs afin de déterminer l'écart de similitude entre votre requête de recherche et tous les vecteurs de caractéristiques disponibles qui représentent des objets détectés.

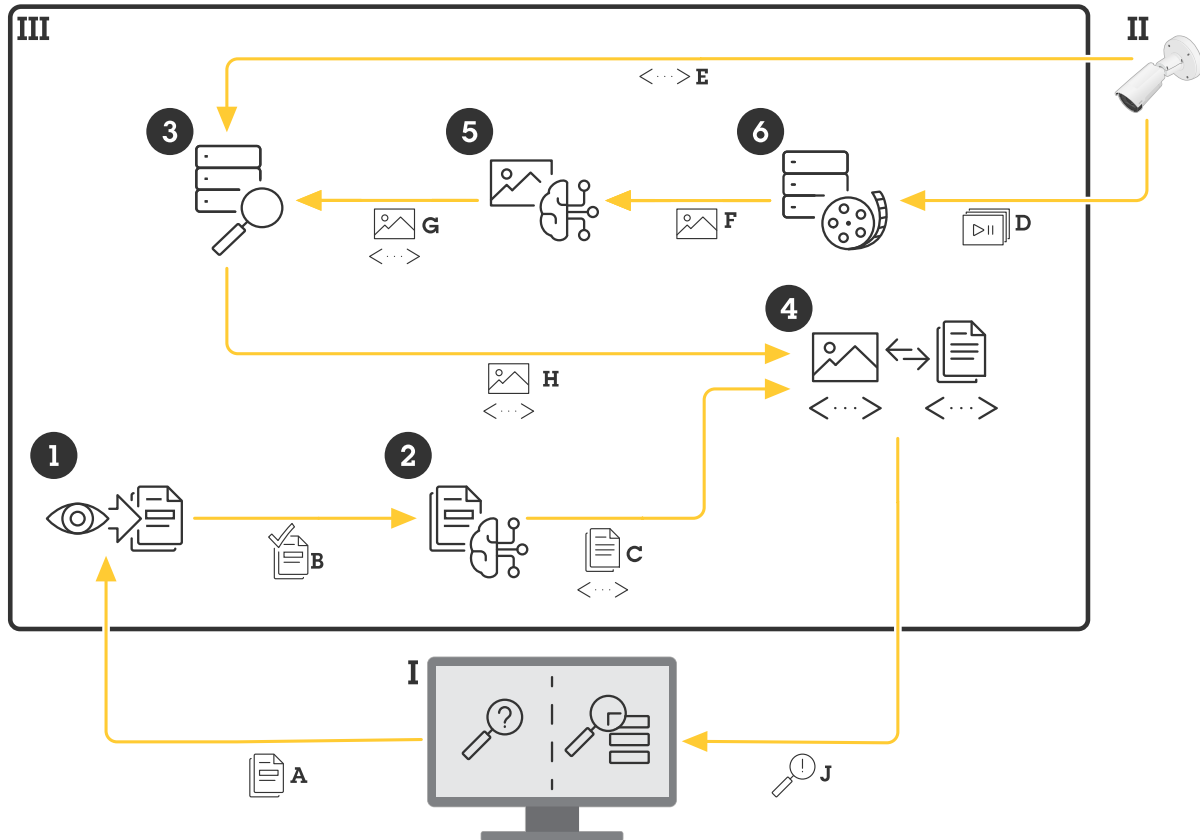
Un vecteur de caractéristiques est une représentation numérique d'un texte ou d'une image. Les vecteurs de caractéristiques de personnes ou d'objets ne sont donc que des représentations abstraites de l'apparence de la personne ou de l'objet. Les vecteurs de caractéristiques ne contiennent pas d'informations humainement interprétables sur des caractéristiques comme la couleur des cheveux ou des vêtements, qui peuvent être explicitement associées à une personne spécifique ou utilisées à des fins d'identification. Les vecteurs de caractéristiques ne sont exploitables que pour les comparer à d'autres vecteurs de caractéristiques.

# 6 Aperçu du processus

Le schéma général du processus présente les principales étapes, y compris l'emplacement où elles se déroulent et le type de données produites à chaque étape.

Notez que la boucle supérieure du diagramme, comprenant la caméra (1), le stockage des enregistrements (6), le modèle de fondation (5) et la base de données de recherche (3), est un processus continu qui

crée des vecteurs de caractéristiques pour tous les objets mobiles détectés, et pas seulement lorsque vous effectuez une recherche.



Principaux emplacements (I-III) du processus de recherche en texte libre

I Client AXIS Camera Station Pro

II Caméra(s)

III Serveur AXIS Camera Station Pro

Principales étapes du processus (1-6)

1 Modération des requêtes de recherche

2 Modèle de fondation (texte)

3 Base de données de recherche

4 Comparaison des vecteurs

5 Modèle de fondation (image)

6 Stockage des enregistrements

Type de données ou résultat (A-J)

A Chaîne de texte

B Chaîne de texte

C Vecteur de caractéristiques (texte)

D Vidéo

- Métadonnées

F Images

G Vecteurs de caractéristiques (image)

H Vecteurs de caractéristiques (image)

J Résultats de la recherche



(I) **Client AXIS Camera Station Pro** : Ici, vous tapez votre requête de recherche et vous obtenez des résultats triés.

(II) **Caméra(s)** : La recherche en texte libre fonctionne sur les caméras Axis exécutant AXIS OS 5.51 ou une version ultérieure, mais plus la caméra est performante, meilleurs sont les résultats obtenus. Les dispositifs plus anciens fournissent des métadonnées moins granulaires basées uniquement sur la détection des mouvements. Les dispositifs plus récents produisent des métadonnées de scène avec AXIS Scene Metadata, qui comprend la classification des objets. La détection et le suivi des objets mobiles par la caméra sont utilisés pour trouver une image représentative de chaque objet détecté, limitant ainsi le nombre d'images à analyser sur le serveur.

(III) **Serveur AXIS Camera Station Pro** : C'est là que sont traitées et stockées toutes les métadonnées et données vidéo provenant des caméras. Avant d'effectuer une recherche en texte libre, le serveur doit décoder la vidéo et extraire une image de chaque objet mobile détecté. Le modèle de fondation traite ensuite cette image pour créer le vecteur de caractéristiques. Ces opérations sont assez intensives en traitement. Pour améliorer les performances, les vecteurs de caractéristiques sont enregistrés dans une base de données pour accélérer les recherches ultérieures. Si votre serveur dispose d'une capacité libre, nous vous recommandons vivement d'activer le traitement en arrière-plan de la vidéo provenant de vos caméras les plus importantes, car cela accélérera considérablement la recherche.

(1) **Modération des requêtes de recherche** : le modèle de modération vérifie la requête pour limiter le contenu offensant.

(2) **Modèle de fondation (texte)** : le modèle de fondation crée une représentation numérique (vecteur de caractéristiques) de la chaîne de texte de la requête de recherche modérée.

(3) **Base de données de recherche** : la base de données de recherche contient les métadonnées complètes provenant d'AXIS Scene Metadata ou des métadonnées créées par le serveur, notamment données de classification des objets avec attributs, heure, position et vecteurs de caractéristiques.

(4) **Comparaison des vecteurs** : la représentation du vecteur de caractéristiques correspondant à la chaîne de texte de la requête de recherche est comparée aux représentations du vecteur de caractéristiques des images d'objet détectées dans la vidéo.

(5) **Modèle de fondation (image)** : le modèle de fondation crée des représentations numériques (vecteurs de caractéristiques) de la trajectoire de chaque objet dans la vidéo enregistrée. Il s'agit d'un processus continu qui se déroule même lorsque vous n'effectuez pas recherche.

(6) **Stockage des enregistrements** : c'est là que la vidéo de la caméra est stockée et que le modèle de fondation obtient ses images.

## 7 Usage responsable de l'IA

Avec la recherche en texte libre, nous utilisons l'IA pour augmenter la précision et l'efficacité de nos solutions de recherche forensique et, au final, améliorer la prise de décision humaine.

La responsabilité et l'obligation de rendre compte sont fondamentales pour l'approche d'Axis en matière d'IA. Cela implique de s'assurer que les systèmes d'IA que nous créons adhèrent aux principes éthiques, respectent les lois et gèrent efficacement les risques. Axis fournit des outils qui permettent à nos clients d'avoir confiance dans l'intégrité de leurs opérations. La fonction de recherche en texte libre d'AXIS Camera Station Pro comprend un modèle de classification de texte affiné pour la modération des invites de texte. Nous avons développé ce modèle pour éviter les abus dans les requêtes de recherche afin de vous aider à prévenir tout usage contraire à l'éthique.

La recherche en texte libre se connecte aux services cloud d'Axis une fois par semaine pour vérifier si les modèles d'IA ont besoin d'être actualisés afin de respecter de nouvelles réglementations ou exigences. En cas d'échec de connexion, les opérations de recherche en texte libre ne seront plus disponibles tant que la connexion n'est pas rétablie.

Pour mieux respecter les normes juridiques et éthiques dans l'application de l'IA, nos produits fournissent des contrôles d'accès basés sur les identifiants et les autorisations d'accès des utilisateurs. Ainsi, nos clients peuvent s'assurer que les utilisateurs se conforment aux politiques opérationnelles.



# À propos d'Axis Communications

Axis contribue à un monde plus sûr et plus intelligent en créant des solutions qui améliorent la sécurité et la performance de l'activité. En tant qu'entreprise leader des technologies réseau, Axis propose des solutions de vidéosurveillance, de contrôle d'accès, de communication vidéo et d'audio. Ces solutions s'accompagnent d'applications d'analyse intelligente et de séances de formation de qualité.

Comptant un effectif d'environ 4000 salariés compétents dans plus de 50 pays, Axis collabore avec des partenaires technologiques et intégrateurs de systèmes pour fournir des solutions aux clients du monde entier. Axis a été fondée en 1984, son siège est situé à Lund en Suède.