

# Analítica de audio para seguridad y prevención

Febrero 2024

## Resumen

Las soluciones de analítica de audio para seguridad pueden detectar patrones de sonido e identificar sonidos inesperados en una secuencia de audio en directo. La analítica, por ejemplo, podría identificar situaciones de violencia verbal para que la situación no vaya a peor, detectar el ruido de cristales rotos para evitar intrusiones o alertar de personas que sufren algún tipo de problema.

El uso de diferentes tipos de sensores, por ejemplo de vídeo y de audio (una cámara y un micrófono), aumenta la fiabilidad de los resultados de la detección y permite obtener información más útil.

AXIS Audio Analytics es una plataforma de software para la analítica en el extremo integrada en el software de los dispositivos Axis. Estas soluciones de analítica capturan y detectan sonidos sin guardar la transmisión de audio original. De este modo, se garantiza la protección de la privacidad, ya que la analítica se ejecuta en el extremo y genera metadatos de audio.

# Índice

1	Introducción	4
2	tecnología	4
2.1	Detección de eventos de sonido	4
2.2	Captura y procesamiento del audio	4
2.3	Analítica en el extremo o basada en servidores	5
2.4	Metadatos	6
3	AXIS Audio Analytics	7
3.1	Privacidad	7
3.2	Aplicaciones	8
4	Restricciones legales	12
5	Nota legal	12

# 1 Introducción

Las soluciones de analítica de audio para seguridad pueden detectar patrones de sonido e identificar sonidos inesperados en una secuencia de audio en directo, como sonidos asociados a agresiones o cristales rotos, o avisar si hay personas que sufren algún tipo de problema.

La analítica de audio, combinada con la videovigilancia, puede alertar a los operadores de posibles incidentes que se están produciendo e indicarles las vistas de cámara correspondientes. De este modo, la detección se produce antes, y también la intervención, lo que en muchos casos ayuda a prevenir males mayores.

Este documento técnico explica cómo puede utilizarse la analítica de audio en el terreno de la seguridad. Hablaremos sobre la tecnología empleada para la captura y el procesamiento del audio, con un especial énfasis en la analítica en el extremo en tiempo real, como las soluciones de clasificación de sonidos basadas en IA y aplicadas directamente en la cámara o el micrófono. También presentaremos las opciones de la analítica en el extremo para proteger la privacidad a través del uso de metadatos de audio.

Este documento no tiene en cuenta los aspectos legales. Por lo tanto, antes de instalar un sistema de vigilancia deberá analizar qué leyes y normativas son de aplicación en su región y en su contexto de uso. El propietario del sistema es el responsable de garantizar el cumplimiento de la legislación, las normativas y las prácticas recomendadas de cada región.

## 2 tecnología

### 2.1 Detección de eventos de sonido

Un evento de sonido es un segmento de audio que las personas identificarían habitualmente como un elemento distintivo, por ejemplo, gritos o cristales rotos. Este tipo de sonidos conceptuales pueden detectarse y etiquetarse de una forma similar a como se detectan y etiquetan las clases de objetos en la analítica de vídeo.

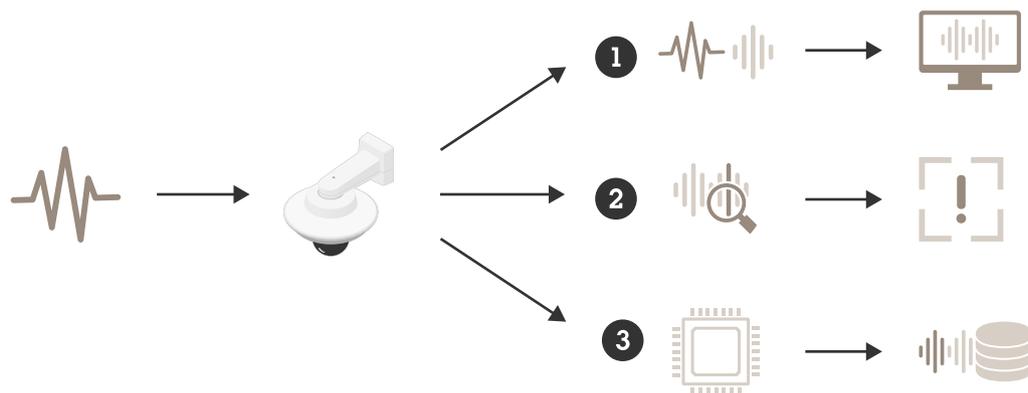
Una solución de analítica entrenada para reconocer patrones de sonido normalmente trata de identificar una combinación de características, que van desde el nivel de decibelios hasta la energía en las diferentes frecuencias a lo largo del tiempo. Cuando se detecta un patrón de sonido concreto, el sistema puede enviar una notificación automática al personal a través de alertas visuales o mediante la activación de una alarma.

Si se utilizan algoritmos basados en IA, pueden entrenarse a partir de grandes volúmenes de datos. Un algoritmo, por ejemplo, puede detectar de forma fiable gritos humanos si se ha entrenado con miles de sonidos de este tipo.

### 2.2 Captura y procesamiento del audio

La analítica de audio utiliza datos de audio capturados y analiza las características relevantes del sonido para generar resultados en un formato distinto al audio. Capturar el audio significa digitalizarlo y ponerlo a disposición del software. Este proceso consiste en capturar las vibraciones de sonido en el aire usando un micrófono, convertir estas señales analógicas en señales digitales y transferirlas a una unidad de procesamiento. Si el audio capturado no se guarda en un soporte permanente, como una memoria flash o un disco duro, no se graba. En las cámaras Axis, la transmisión y la grabación de audio están desactivadas de forma predeterminada.

Una vez capturado el audio, la información capturada se prepara para los siguientes pasos de procesamiento. Los diferentes preparativos pueden realizarse en paralelo o de forma independiente.



#### Preparación del audio

- 1 Transformación
- 2 Análisis en el extremo en tiempo real
- 3 Procesamiento y codificación para transmisión o almacenamiento: si utiliza una cámara Axis, el audio no se transmite ni almacena si no activa la transmisión de audio.

- **Transformación.** El sonido pasa a ser abstracto y se convierte, por ejemplo, en información visual, como un gráfico que muestra el espectro acústico. No es posible volver al punto de origen: es imposible recuperar el sonido original a través del gráfico del espectro.
- **Análisis en el extremo en tiempo real.**

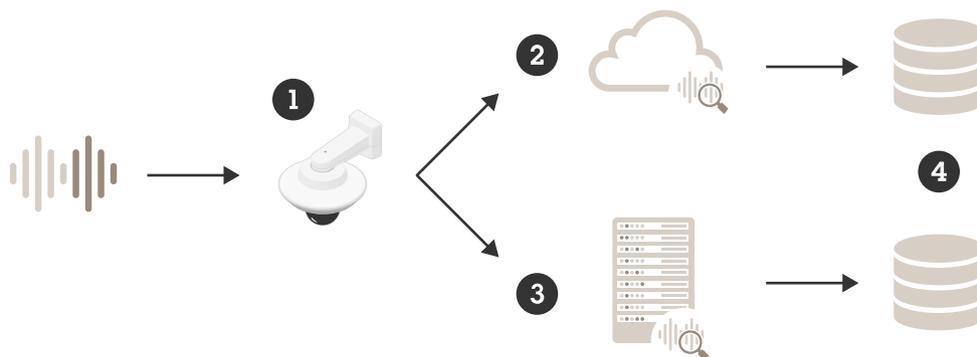
Se puede utilizar un **clasificador de sonido** si el audio se procesa en el extremo. El resultado son metadatos que describen las características del sonido. No es posible recrear el sonido original a partir de sus metadatos.

Se puede utilizar un **detector de sonido** para reconocer patrones, niveles o frecuencias y proporcionar información sobre el estado. De nuevo, es imposible recuperar el sonido original.
- **Procesamiento y codificación.** En aquellos casos en los que se utilizará el audio original (sin transformar ni analizar), suele incluirse una parte de procesamiento y codificación para preparar los datos de audio para una finalidad concreta. Algunas opciones serían guardar los datos de audio en el extremo, transmitirlos a clientes externos para su procesamiento (en un servidor o en la nube) o almacenarlos en servicios externos. Con una cámara Axis, en primer lugar debe activar la transmisión de audio, que está desactivada de forma predeterminada por motivos de privacidad (control de privacidad del audio).

### 2.3 Análisis en el extremo o basada en servidores

El lugar del sistema donde se encuentra el motor de la analítica es importante por muchas razones. Sobre todo para las cuestiones relacionadas con la privacidad y el cumplimiento de los reglamentos sobre información personal, es determinante dónde analiza los datos de audio el algoritmo de software. Hay situaciones en las que los datos de audio no pueden transmitirse por la red y es vital que los datos de audio capturados (pero no guardados) puedan analizarse en local. Sin embargo, cuando se necesitan algoritmos

con una gran potencia de computación, seguramente no podrán ejecutarse en el extremo y deberán enviarse los datos de audio digital a la nube o a un servidor.



#### *Posibles puntos de aplicación de la analítica*

- 1 Extremo
- 2 Nube
- 3 Servidor
- 4 Almacenamiento: solo puede almacenarse el audio original si la transmisión de audio está activada.

- **Analítica en el extremo.** Cuando la analítica se ejecuta en el extremo, la transmisión de audio no sale de la cámara. Solo lo hace la información de la analítica aplicada, esto es, metadatos o activadores. En AXIS Audio Analytics se utiliza analítica en el extremo.
- **Analítica en servidores.** Cuando la analítica se ejecuta en un servidor, la cámara debe transmitir datos de audio. Si se preprocesan en la cámara, tal vez solo puedan obtenerse metadatos sintetizados o despersonalizados. Un servidor normalmente forma parte de un sistema cerrado (el control está en manos del propietario del sistema), por lo que es posible gestionar los aspectos relativos a la privacidad del audio transmitido. De todos modos, es importante respetar las regulaciones y normativas aplicables.
- **Analítica en la nube.** Los datos de audio también pueden transmitirse a un servidor en una nube. Al igual que sucede con la analítica en un servidor, la información de audio puede procesarse antes y convertirse en metadatos. Como la utilización de la nube suele estar descentralizada, es todavía más importante abordar las cuestiones de privacidad y garantizar el cumplimiento de las normativas.

## 2.4 Metadatos

La analítica de audio genera un flujo constante de metadatos a partir de datos de nivel de audio. La analítica también genera metadatos a partir de eventos detectados por clasificadores y mediante la detección de audio adaptativa. La analítica en el extremo analiza la información de audio en la cámara. No hace falta transmitir el flujo de audio real a ninguna parte: basta con transmitir solo los metadatos, que permiten saber qué ocurre en la escena.

El flujo de metadatos permite una inspección visual del envolvente acústico, por ejemplo representado en un panel en el que los eventos de audio y vídeo pueden combinarse y visualizarse juntos. De este modo, los metadatos ayudan a buscar de forma rápida y eficaz eventos concretos y sonidos inesperados, lo que ahorra a los investigadores horas y horas de trabajo analizando ingentes cantidades de vídeo.

## 3 AXIS Audio Analytics

AXIS Audio Analytics es una plataforma de software que continuamente introduce soluciones y funciones de analítica nuevas y mejoradas. Las soluciones de analítica están integradas en el software AXIS OS del dispositivo y se incluyen gratis con las cámaras y otros dispositivos Axis.

Como funcionan en el extremo, los algoritmos se aplican directamente en la cámara, lo que garantiza una escalabilidad óptima, un tráfico de datos reducido y la máxima privacidad. Solo se guardan los resultados de la analítica (metadatos o activadores). No se graba ni transmite audio de la cámara y no es posible recrear los sonidos originales a partir de los metadatos.

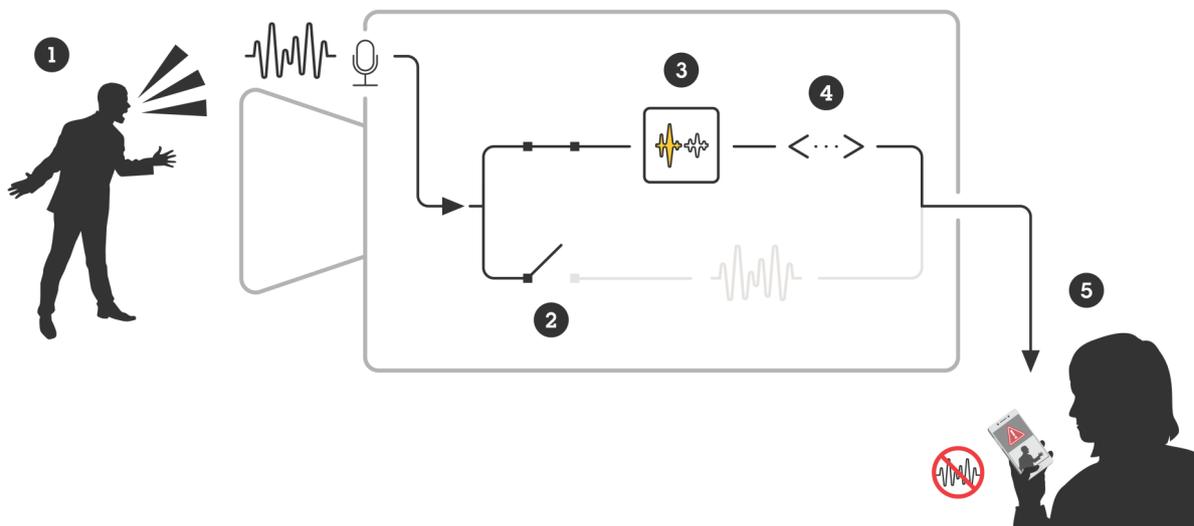
- **Clasificación de audio (gritos y chillidos).** Este clasificador de sonido basado en IA detecta y analiza sonidos específicos, como gritos y chillidos. En el futuro se añadirán más clasificaciones de audio, por ejemplo, sonido de cristales rotos. El resultado de la analítica de clasificación son unos metadatos que describen las características del sonido.
- **Detección de audio adaptativa.** Se trata de un detector de sonido que crea un evento cuando se produce un cambio repentino en el nivel de audio. Puede detectar picos de sonido de cualquier tipo y, además, se adapta al ruido ambiental aunque varíe el nivel de ruido.

### 3.1 Privacidad

En general, la analítica de audio no graba el audio entrante ni lo transmite desde la cámara. Simplemente procesa sonidos para permitir la búsqueda de eventos específicos, patrones o niveles de sonido en un sistema receptor, como un panel para profundizar en los análisis o un software de gestión de vídeo para alertar a operadores. No es posible reconstruir los datos de audio ni grabar conversaciones privadas. Y esto es así porque la analítica se ejecuta en el extremo y genera metadatos de audio.

AXIS Audio Analytics no graba ni transmite audio, sino que únicamente transmite metadatos. Por motivos de privacidad, las transmisiones de audio están desactivadas de forma predeterminada en los dispositivos Axis (control de privacidad del audio), lo que implica que el audio no puede transmitirse, grabarse ni reconstruirse. Puede activar la transmisión de audio si resulta necesario, aunque recibirá una notificación cuando se produzca algún hecho relevante, aunque la transmisión de audio esté desactivada.

La ilustración muestra la interacción de AXIS Audio Analytics con el control de privacidad del audio para capturar sonidos y utilizar metadatos para crear una alerta.



*Control de privacidad de audio y AXIS Audio Analytics en las cámaras Axis, de un grito (1) a una alerta (5).*

- 1 El micrófono captura los sonidos.*
- 2 La transmisión de audio está desactivada de forma predeterminada.*
- 3 El clasificador de audio de AXIS Audio Analytics detecta los gritos o chillidos en el audio entrante.*
- 4 El clasificador de audio genera metadatos, que incluyen una notificación de evento.*
- 5 Las personas interesadas reciben una alerta basada en la notificación y los metadatos del evento. Pueden verificarla revisando la transmisión de vídeo. No hay ninguna transmisión de audio disponible.*

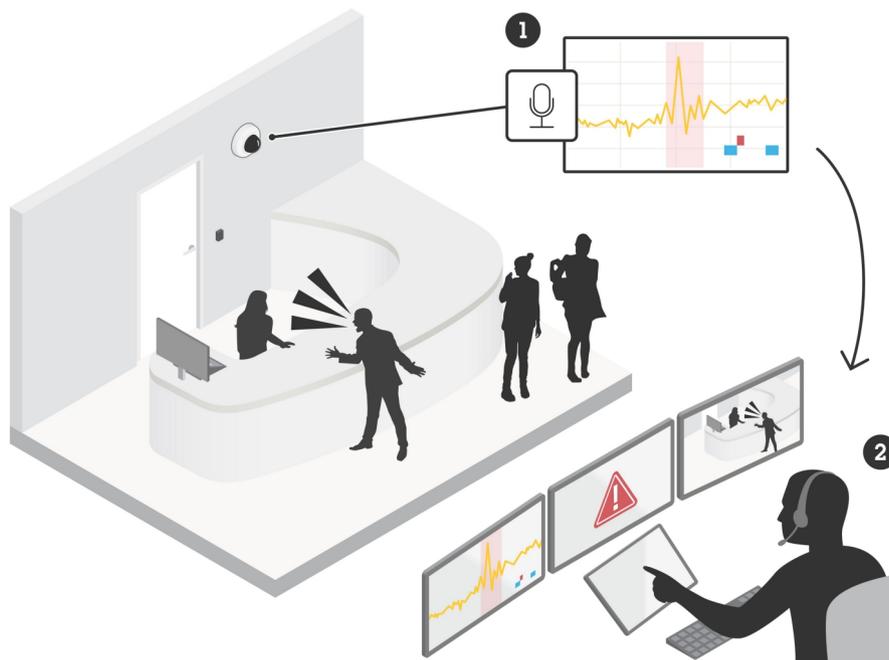
## 3.2 Aplicaciones

Aunque la analítica basada en IA atesora un gran potencial a la hora de eliminar ruidos sin interés, también puede generar falsas alertas cuando hay mucho ruido de fondo. La lluvia contra las ventanas, truenos, sirenas, música o lugares con mucha gente hablando pueden generar falsas alertas. Por lo tanto, los entornos de uso más habituales son espacios tranquilos como bancos o zonas de recepción, pero también otros espacios interiores fuera del horario laboral, como tiendas, restaurantes, escaleras u oficinas.

### 3.2.1 Detección de sonido con alertas

En un banco o una recepción, una solución de analítica con clasificación de sonidos puede analizar el espacio y detectar sonidos asociados a agresiones o cristales rotos. En el momento de la detección, los metadatos de audio y los datos de eventos del sistema avisan automáticamente al personal a través de una

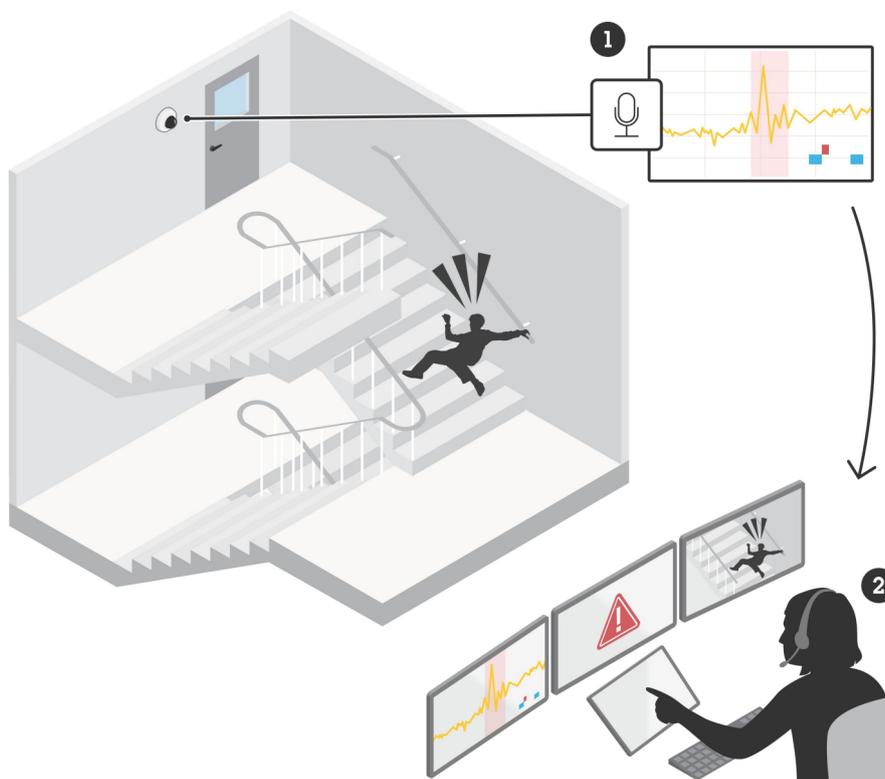
alerta visual o mediante la activación de una alarma. Este sistema facilita una detección precoz y, por lo tanto, permite responder e intervenir con mayor rapidez.



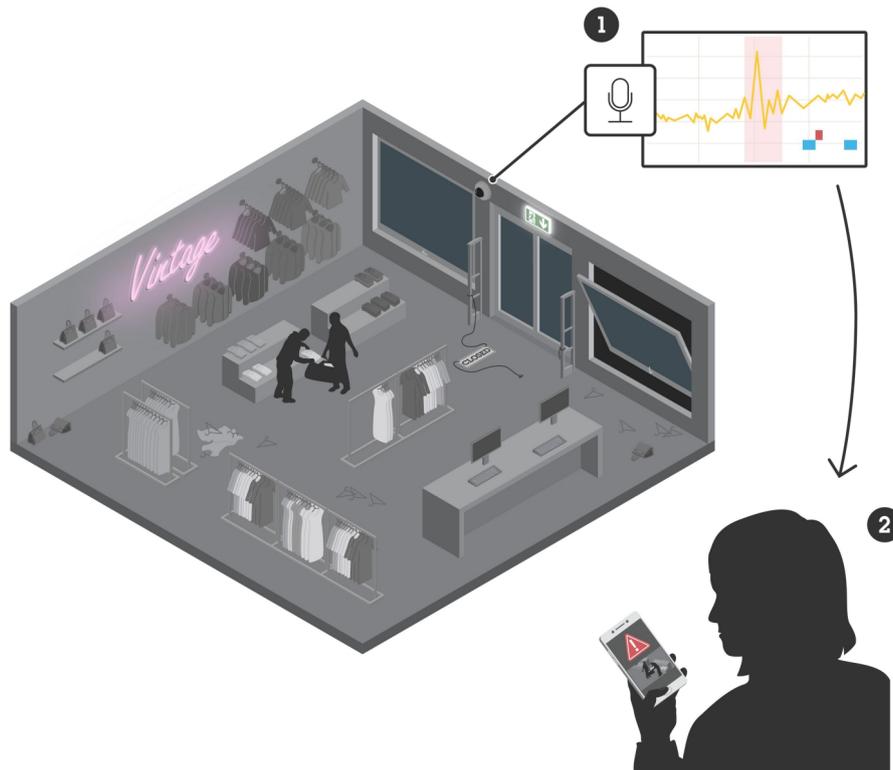
- 1 Una cámara con analítica de clasificación de sonidos detecta gritos o una discusión en la recepción.
- 2 Un operador recibe una alerta, que puede verificar revisando la transmisión de vídeo antes de tomar medidas.

La analítica de detección de audio adaptativa puede utilizarse para detectar sonidos inesperados fuera del horario laboral. La analítica analiza el sonido ambiente y responde si detecta voces, ventanas rotas u

otros ruidos repentinos y breves. Cuando se detecta un evento, la analítica transmite los metadatos para informar a los operadores de la situación.



- 1 Una cámara con analítica de clasificación de sonidos detecta sonidos inesperados fuera del horario laboral.
- 2 Un operador recibe una alerta, que puede verificar revisando la transmisión de vídeo antes de tomar medidas.



- 1 Una cámara con analítica de clasificación de sonidos detecta sonidos en una tienda fuera del horario de apertura.
- 2 El propietario de la tienda recibe una alerta, que puede verificar revisando la transmisión de vídeo antes de tomar medidas.

### 3.2.2 Combinación de sensores para sacar más partido al sistema de vigilancia

Los sistemas de vigilancia a menudo incorporan varios tipos de sensores. Uno de ellos es el sensor de imagen de la cámara, responsable de capturar el componente visual de una escena. Son también habituales los sensores no visuales, como los detectores de movimiento basados en la tecnología de radar o en emisiones de radiación infrarroja. Los sensores no visuales complementan la instalación de cámaras aportando otros tipos de información.

Además, cuando se añaden sensores de audio (micrófonos) a una instalación de vigilancia, hay ventajas para la gran mayoría de casos de uso potenciales. Al sumar funciones de audio y de analítica de audio a un sistema que no es de audio se abre la puerta a la interacción entre múltiples sensores. Si utiliza analítica de vídeo, al combinarla con analítica de audio mejorará la fiabilidad de la detección, especialmente en situaciones en las que la analítica de vídeo pierda eficacia a causa de situaciones de poca luz.

Tiene la posibilidad de configurar el sistema, por ejemplo en el software de gestión de vídeo, para que active acciones solo si reaccionan tanto la analítica de vídeo como la analítica de audio. Por ejemplo, cuando la analítica de audio detecta un grito y la analítica de vídeo detecta una persona dentro del campo de visión de la cámara. En algunos entornos, esta combinación garantiza el nivel de seguridad adecuado.

### 3.2.3 Presentación de datos en paneles

Los metadatos de audio pueden transferirse a paneles de analítica, por ejemplo plataformas de VSaaS (videovigilancia como servicio), IoT (internet of things) o de inteligencia empresarial, que recopilan los metadatos y los presentan de forma visual. Estos paneles analizan tendencias actuales e históricas para obtener un resumen de la situación y también información útil para adoptar medidas. Los análisis estadísticos basados en la afluencia o la experiencia del cliente ayudan a tomar decisiones basadas en datos con el objetivo de mejorar las operaciones.

Los paneles permiten ver resultados sin necesidad de escuchar el audio en sí o recuperar el sonido original. La información clave puede obtenerse, por ejemplo, a partir de eventos de conteo, y no puede haber dudas en relación con la privacidad de los datos. Tenga en cuenta que tal vez existan diferencias en las restricciones legales en función de si el audio se graba o solo se captura.

## 4 Restricciones legales

A mucha gente le preocupa el uso del micrófono en contextos de videovigilancia. Normalmente lo que les preocupa es que se graben conversaciones junto con el material de vídeo. Sin embargo, con la analítica de audio normalmente no se graba ni se transmite audio. Las leyes que regulan la vigilancia son diferentes en cada región y en cada país, por lo que es importante investigar qué está permitido antes de usar el audio en un sistema de vigilancia.

La captura y la grabación del audio pueden estar prohibidas o sujetas a consideraciones especiales por varios motivos, según lo establecido por la legislación nacional o por varios tipos de normativas y reglamentos locales. Una misma región o espacio puede permitir la captura de audio pero prohibir las grabaciones. Las empresas pueden prohibir también la vigilancia de audio en sus instalaciones.

## 5 Nota legal

Este documento y su contenido son cortesía de Axis, y todos los derechos sobre el documento y los derechos de propiedad intelectual derivados del mismo (lo que incluye, a título meramente ilustrativo, marcas comerciales, nombres de marcas, logotipos y marcas similares) están protegidos por la ley. Asimismo, todos los derechos y/o intereses vinculados al documento o derechos de propiedad intelectual relacionados con el mismo son y serán propiedad de Axis Communications AB.

Este documento se proporciona "tal cual", sin garantía de ningún tipo y únicamente con una finalidad informativa. La información facilitada en este documento no pretende ofrecer ningún tipo de asesoramiento legal. Asimismo, no pretende crear ningún tipo de obligación legal para Axis Communications AB ni para sus filiales. Las obligaciones de Axis Communications AB y/o sus filiales en relación con cualquier producto Axis están sujetas exclusivamente a los términos y condiciones del contrato entre Axis y la persona física o jurídica que compra dicho producto a Axis directamente.

PARA EVITAR DUDAS, TODO EL RIESGO EN CUANTO AL USO, RESULTADOS Y RENDIMIENTO DE ESTE DOCUMENTO ES ASUMIDO POR EL USUARIO DEL MISMO Y AXIS RECHAZA Y EXCLUYE, EN LA MEDIDA MÁXIMA PERMITIDA POR LA LEY, TODAS LAS GARANTÍAS, YA SEAN LEGALES, EXPRESAS O IMPLÍCITAS, INCLUYENDO, ENTRE OTRAS, CUALQUIER GARANTÍA IMPLÍCITA DE COMERCIALIZACIÓN, IDONEIDAD PARA UN PROPÓSITO PARTICULAR, TITULARIDAD Y NO INFRACCIÓN Y RESPONSABILIDAD DEL PRODUCTO, O CUALQUIER GARANTÍA QUE SURJA DE CUALQUIER PROPUESTA, ESPECIFICACIÓN O MUESTRA CON RESPECTO A ESTE DOCUMENTO.



# Acerca de Axis Communications

Axis contribuye a crear un mundo más inteligente y seguro a través de soluciones para mejorar la seguridad y el rendimiento empresarial. Como empresa de tecnología de red y líder del sector, Axis ofrece soluciones de videovigilancia, control de acceso y sistemas de audio e intercomunicación. Se ven reforzadas por aplicaciones de análisis inteligentes y respaldadas por formación de alta calidad.

Axis tiene alrededor de 4000 empleados dedicados en más de 50 países y colabora con socios de integración de sistemas y tecnología en todo el mundo para ofrecer soluciones personalizadas. Axis se fundó en 1984 y la sede está en Lund, Suecia