

WHITE PAPER

# Analíticos de áudio para proteção e segurança

Fevereiro 2024

## Resumo

Os analíticos de áudio para segurança e proteção podem detectar padrões de som e destacar sons inesperados em áudios ao vivo. Os analíticos podem, por exemplo, identificar uma agressão verbal para evitar o agravamento do conflito e as possíveis agressões, detectar quebras de vidros para impedir arrombamentos ou emitir avisos precoces de pessoas em perigo.

O uso de diferentes tipos de sensores, como vídeo e áudio (uma câmera e um microfone), aumenta a confiança nos resultados de detecção e contribui para informações mais acionáveis.

O AXIS Audio Analytics é uma plataforma de software para analíticos baseados na borda, integrada ao software dos dispositivos Axis. Esses analíticos capturam e detectam sons sem guardar o stream de áudio original. Essa é uma forma de proteger a privacidade e funciona bem porque os analíticos se baseiam na borda e fornecem metadados de áudio.

# Sumário

1	Introdução	4
2	Tecnologia	4
	2.1 Detecção de eventos sonoros	4
	2.2 Captura e processamento do áudio	4
	2.3 Analíticos baseados na borda ou em servidor.	5
	2.4 Metadados	6
3	AXIS Audio Analytics	7
	3.1 Privacidade	7
	3.2 Casos de uso	8
4	Restrições legais	12
5	Aviso de isenção de responsabilidade	12

# 1 Introdução

Os analíticos de áudio para segurança e proteção podem detectar padrões de som e destacar sons inesperados em áudios ao vivo. Os analíticos podem, por exemplo, identificar sons associados a agressões, detectar quebras de vidros ou fornecer avisos antecipados de pessoas em perigo.

Os analíticos de áudio combinados com o videomonitoramento podem alertar os operadores sobre possíveis incidentes em andamento e orientá-los para as visualizações de câmera relevantes. Isso permite a detecção precoce, a intervenção rápida e, em muitos casos, a prevenção do agravamento da situação.

Este white paper apresenta como os analíticos de áudio podem ser usados para segurança e proteção. Discutimos a tecnologia de captura e processamento de áudio, com foco nos analíticos de borda em tempo real, como classificação de som baseada em IA diretamente na câmera ou microfone. Também mostramos como os analíticos de áudio na borda oferecem diversas opções para manter a privacidade por meio do uso de metadados de áudio.

Este artigo não fornece aconselhamento jurídico. Antes de instalar qualquer sistema de monitoramento, você precisa investigar quais leis e regulamentos se aplicam à sua região e ao seu caso de uso. Cabe ao proprietário do sistema garantir que esteja em conformidade com as leis, regulamentos e recomendações regionais.

## 2 Tecnologia

### 2.1 Detecção de eventos sonoros

Um evento sonoro é um segmento de áudio que os humanos geralmente identificariam como um conceito distinto, por exemplo, o conceito de grito ou vidro quebrando. Esses tipos de sons conceituais podem ser detectados e classificados de maneira semelhante às classes de objetos detectadas e classificadas nos analíticos de vídeos.

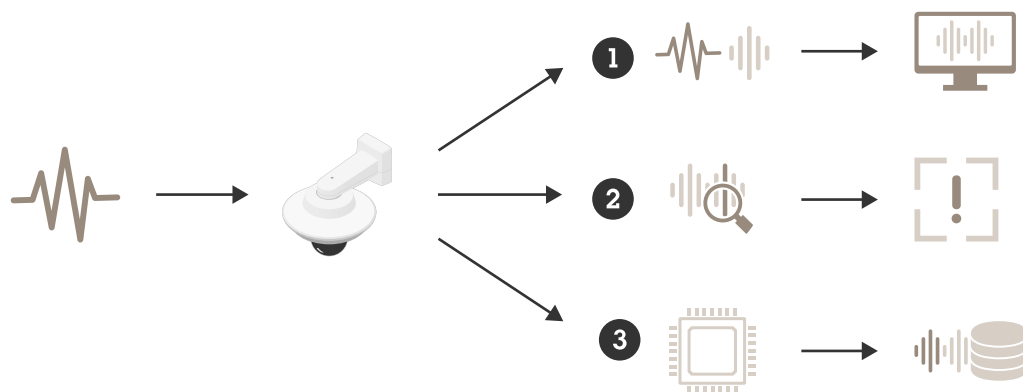
Os analíticos que são treinados para reconhecer padrões de som costumam captar uma combinação de características que podem englobar do nível de decibéis à energia em diferentes frequências ao longo do tempo. Quando um padrão sonoro específico é detectado, o sistema pode enviar uma notificação automática à equipe por meio de um alerta visual ou acionando um alarme.

Se você usar algoritmos baseados em IA, eles poderão ser treinados usando grandes quantidades de dados. Por exemplo, um algoritmo pode detectar com segurança gritos humanos depois de ter sido treinado ouvindo milhares desse tipo de som.

### 2.2 Captura e processamento do áudio

Os analíticos de áudio usam dados do áudio capturado e analisam características relevantes do som para gerar uma saída que não é áudio. Capturar áudio basicamente significa digitalizá-lo e torná-lo disponível para uso em software. Isso é feito capturando as vibrações sonoras no ar usando um microfone, convertendo esses sinais analógicos em digitais e passando-os para a unidade de processamento. Se o áudio capturado não for colocado em uma mídia permanente, como uma memória flash ou um disco rígido, não é gravado. Nas câmeras Axis, os streams e as gravações de áudio ficam desativados por padrão.

Após a captura de áudio inicial, as informações capturadas são preparadas para as próximas etapas de processamento. As diferentes preparações podem ser feitas em paralelo ou de forma exclusiva.



### Preparação de áudio

- 1 *Transformação*
- 2 *Analíticos de borda em tempo real*
- 3 *Processamento e codificação para streams ou armazenamento: se você usar uma câmera Axis, o áudio não será transmitido nem armazenado, a menos que você ative os streams de áudio.*

- **Transformação** O som é sintetizado e convertido em, por exemplo, informações visuais como um gráfico, mostrando o espectro do som. Este processo não pode ser revertido. Você não pode recuperar o som original do gráfico do espectro.

- **Analíticos de borda em tempo real.**

Um **classificador de som** pode ser usado se o som for processado na borda. O resultado serão metadados que descrevem as características do som. O som original não pode ser recriado a partir de seus metadados.

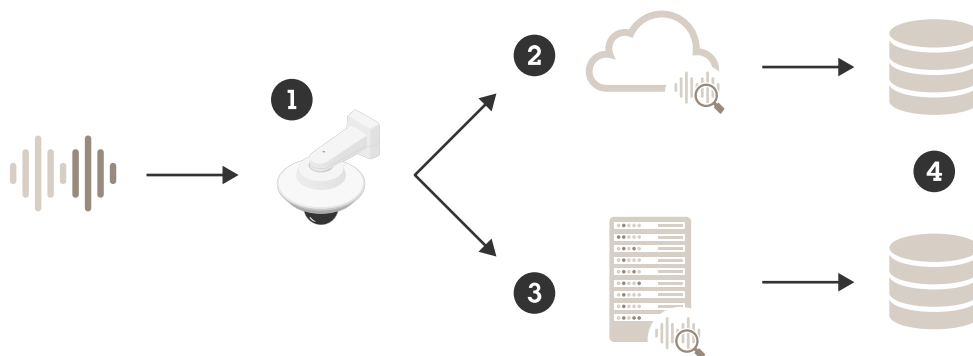
Um **detector de som** pode ser usado para reconhecer padrões, níveis ou frequência e fornecer informações de status. De novo, o som original não é restaurável.

- **Processamento e codificação.** Para os casos em que o áudio original será usado (não transformado ou analisado), algum processamento e codificação é normalmente executado para preparar os dados de áudio para os casos de uso pretendidos. Estes casos de uso podem envolver o armazenamento dos dados de áudio na borda, transmissão para clientes externos para processamento adicional (no servidor ou na nuvem) ou armazenamento externo. Com uma câmera Axis, primeiro você precisa ativar os streams de áudio, que ficam desativados por padrão por motivos de privacidade (controle de privacidade do áudio).

## 2.3 Analíticos baseados na borda ou em servidor.

A localização do mecanismo dos analíticos no sistema é importante por muitos motivos. Especialmente para gerenciar questões de privacidade e cumprir com os regulamentos relacionados aos dados pessoais, é importante onde o algoritmo do software analisa os dados de áudio. Há situações em que os dados de áudio não podem ser enviados pela rede e é fundamental que os dados do áudio capturado (mas não armazenados) possam ser analisados localmente. Se forem necessários algoritmos de computação

intensiva, tais como os que não podem executados na borda, pode ser necessário enviar dados de áudio digital para a nuvem ou para um servidor.



*Possíveis locais para executar os analíticos*

- 1 Edge
- 2 Nuvem
- 3 Servidor
- 4 Armazenamento: o áudio original poderá ser armazenado somente se os streams de áudio estiverem ativados.

- **Analíticos de borda.** Quando os analíticos são executados na borda, nenhum stream de áudio precisa sair da câmera. Somente a saída dos analíticos executados, ou seja, os metadados ou disparos precisam. Os analíticos do AXIS Audio Analytics baseiam-se na borda.
- **Analíticos de servidor.** Quando os analíticos são executados em um servidor, os dados de áudio precisam ser transmitidos da câmera. Se forem pré-processados na câmera, pode resultar em somente metadados abstraídos ou despersonalizados. O servidor geralmente faz parte de um sistema fechado (o proprietário do sistema está no controle), portanto, é possível gerenciar a privacidade do áudio transmitido. No entanto, deve-se garantir que as regras e regulamentos aplicáveis sejam seguidos.
- **Analíticos de nuvem.** Os dados do áudio também podem ser transmitidos para um servidor em um contexto de nuvem. Assim como nos analíticos do servidor, as informações do áudio podem ser pré-processadas em metadados. O uso da nuvem geralmente é descentralizado, por isso é ainda mais importante abordar as questões de privacidade e assegurar que os regulamentos sejam cumpridos.

## 2.4 Metadados

Os analíticos de áudio geram um stream constante de metadados de dados do nível do áudio. Os analíticos também geram metadados com base em eventos detectados por classificadores e detecções adaptativas de áudio. Os analíticos executados na borda avaliam as informações do áudio dentro da câmera. Eles não precisam transmitir o stream de áudio. É possível transmitir somente os metadados, que fornecem informações sobre o que está acontecendo na cena.

O stream de metadados permite a inspeção visual do envelope de áudio, por exemplo, representado em um painel em que os eventos de áudio e os eventos de vídeo podem ser alinhados e visualizados juntos. Desta forma, os metadados permitem pesquisar de forma eficiente e rápida, eventos específicos e sons inesperados. Isso pode economizar muitas horas de pesquisa dos investigadores em altos volumes de filmagens.

## 3 AXIS Audio Analytics

O AXIS Audio Analytics é uma plataforma de software que apresenta continuamente analíticos e recursos novos e atualizados. Os analíticos são integrados ao sistema operacional AXIS OS e incluídos gratuitamente em câmeras e outros dispositivos Axis.

Os analíticos dentro da plataforma baseiam-se na borda, e os algoritmos são executados diretamente na câmera. Isso fornece escalabilidade ideal, baixo tráfego de dados e privacidade. Somente a saída dos analíticos (metadados ou disparos) é guardada. Nenhum áudio é gravado ou transmitido pela câmera, e os sons originais não podem ser recriados usando os metadados.

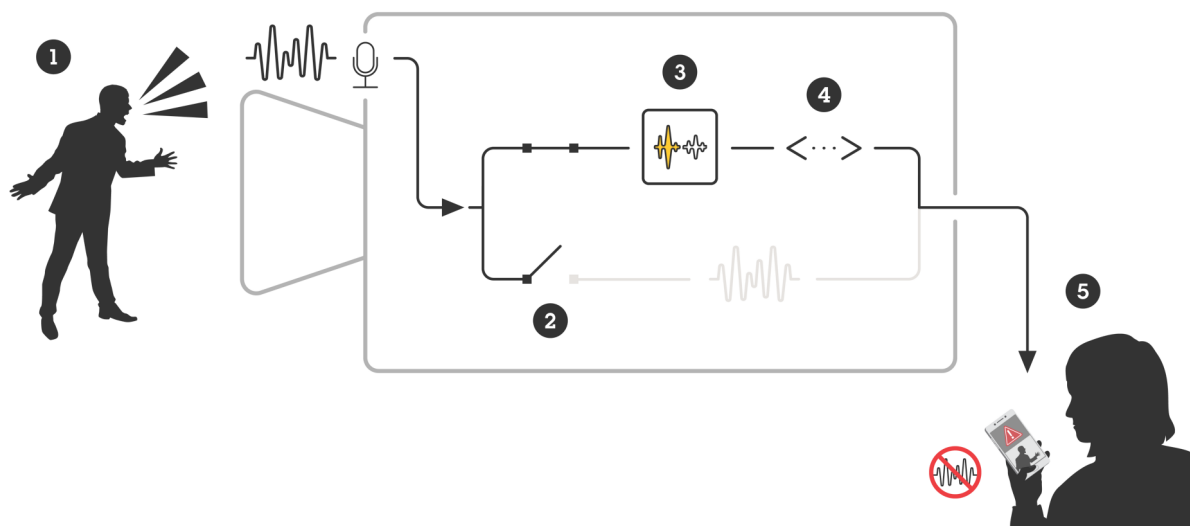
- **Classificação de áudio (grito e berro).** Esse é um classificador de som baseado em IA que detecta e analisa sons específicos, como gritos e berros. Mais classificações de áudio, como quebra de vidro, serão adicionadas no futuro. O resultado dos analíticos de classificação são metadados que descrevem as características do som.
- **Detecção de áudio adaptativa.** Esse é um detector de som que cria um evento quando ocorre uma mudança repentina no nível do áudio. Ele detecta picos sonoros de todo tipo, com a vantagem de se adaptar ao ruído ambiente mesmo quando o nível de ruído varia.

### 3.1 Privacidade

Os analíticos de áudio em geral não gravam o áudio recebido nem o transmitem da câmera. Eles apenas processam os sons para permitir a busca por eventos, padrões ou níveis de som específicos em um sistema receptor, como um painel para investigação detalhada ou um software de gerenciamento de vídeo para alertar os operadores. Nenhum dado do áudio pode ser reconstruído e nenhuma conversa privada pode ser gravada. Isso ocorre porque os analíticos se baseiam na borda e fornecem metadados de áudio.

O AXIS Audio Analytics não grava nem transmite áudio, transmite apenas metadados. Por motivos de privacidade, todos os streams de áudio também ficam desativados por padrão nos dispositivos Axis (controle de privacidade de áudio), o que significa que o áudio não é transmitido e nem gravado, e não é possível recriá-lo. Você pode ativar o stream de áudio se precisar, mas mesmo que ele esteja desativado, você será notificado quando algo relevante acontecer.

A ilustração mostra uma visão geral de como o AXIS Audio Analytics funciona em conjunto com o controle de privacidade de áudio na captação de sons e no uso de metadados para criar um alerta.



*Controle de privacidade de áudio e AXIS Audio Analytics em câmeras Axis desde um grito (1) até o alerta (5).*

- 1 Os sons são captados pelo microfone.*
- 2 Por padrão, o stream de áudio fica desativado.*
- 3 O classificador de áudio do AXIS Audio Analytics detecta gritos ou berros no áudio recebido.*
- 4 Os metadados, incluindo uma notificação do evento, são gerados pelo classificador de áudio.*
- 5 As partes interessadas recebem um alerta com base na notificação do evento e nos metadados. Elas podem verificar ao examinar o stream de vídeo. Nenhum stream de áudio fica disponível.*

## 3.2 Casos de uso

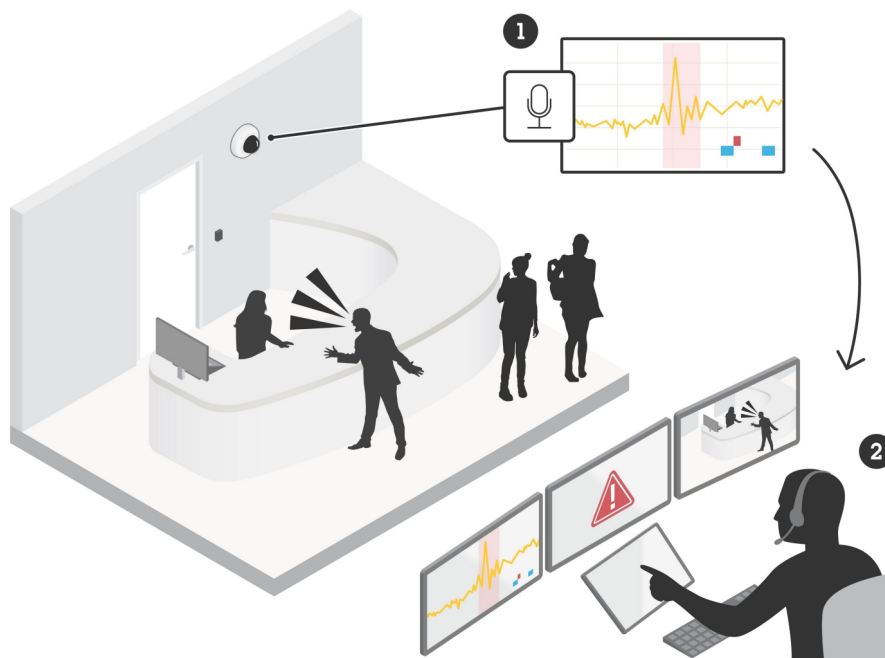
Embora os analíticos baseados em IA consigam filtrar muitos ruídos irrelevantes, eles podem emitir alertas falsos quando há muito ruído de fundo. Chuva batendo nas janelas, trovões, sirenes, música ou cenas movimentadas com pessoas conversando podem acionar alertas falsos. Assim, os casos de uso comuns incluem áreas tranquilas, como bancos e salas de recepção, mas também muitos tipos de espaços internos após o horário comercial, como lojas, restaurantes, escadas ou escritórios.

### 3.2.1 Detecção de som com alertas

Em um banco ou uma sala de recepção, os analíticos de classificação do som podem monitorar a área e detectar sons associados a agressões ou quebra de vidros. Após a detecção, os dados dos eventos e os



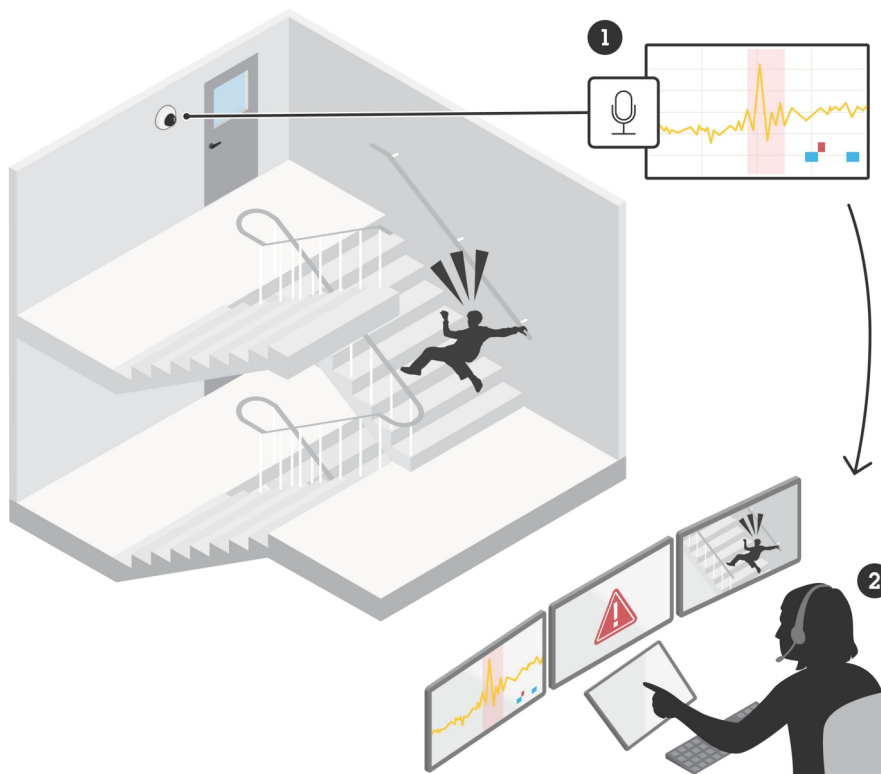
metadados do áudio do sistema enviam notificações automáticas à equipe por meio de um alerta visual ou acionando um alarme. Isso fornece um aviso prévio que permite respostas e intervenção rápidas.



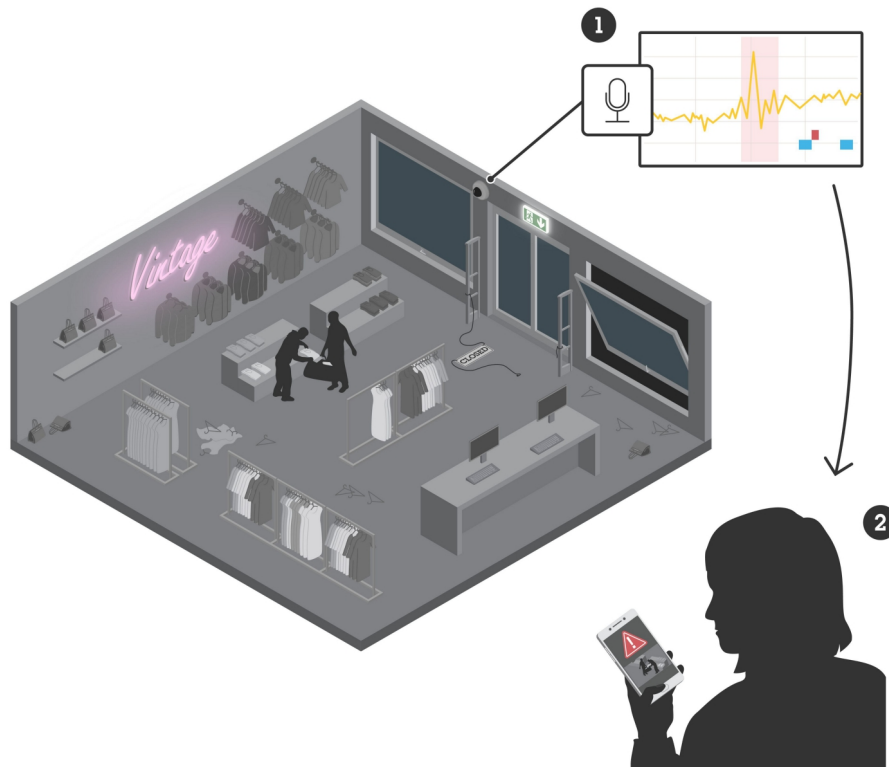
- 1 A câmera com os analíticos de classificação de som detecta gritos ou berros na sala de recepção.
- 2 O operador recebe um alerta e pode verificar o stream de vídeo antes de tomar uma ação.

Os analíticos adaptativos de detecção de áudio podem ser usados para detectar sons inesperados fora do horário comercial. Os analíticos avaliam os sons ambientes e reagem quando detectam vozes, janelas

quebrando ou outros ruídos repentinos e de curta duração. Quando os eventos são detectados, os analíticos encaminham os metadados para notificar os operadores corretamente.



- 1 A câmera com analíticos de classificação de som detectam sons inesperados após o horário comercial.
- 2 O operador recebe um alerta e pode verificar o stream de vídeo antes de tomar uma ação.



- 1 A câmera com analíticos de classificação de som detectam sons em uma loja após o horário comercial.
- 2 O proprietário da loja recebe um alerta e pode verificar o stream de vídeo antes de tomar uma ação.

### 3.2.2 Combinação de sensores para aproveitar melhor o sistema de monitoramento

Os sistemas de monitoramento frequentemente incorporam vários tipos de sensores. O sensor de imagem da câmera é um deles, é claro, registrando o aspecto visual de uma cena. Sensores não visuais também são comumente usados, como detectores de movimento com base em tecnologia de radar ou emissões de radiação infravermelha. Sensores não visuais complementam a instalação da câmera ao adicionar outros tipos de entrada de informações.

Também empregando sensores de áudio (microfones) em uma instalação de monitoramento, a grande maioria de todos os casos de uso possíveis são reforçados. A adição da capacidade de áudio e dos analíticos de áudio a um sistema que não é de áudio permite a interação multisensor. Se estiver usando analíticos de vídeo, a adição dos analíticos de áudio poderá aumentar a confiança da detecção. Esse é o caso especialmente se os analíticos de vídeo forem dificultados, por exemplo, por condições de pouca luz.

Você pode configurar o sistema, por exemplo, em um software de gerenciamento de vídeo, para que ele acione ações somente quando tanto os analíticos de vídeo quanto os analíticos de áudio reagirem. Por exemplo, os analíticos de áudio detectam um grito e os analíticos de vídeo detectam um indivíduo no campo de visão da câmera. Em alguns ambientes, essa combinação proporciona o nível correto de segurança.

### 3.2.3 Fornecimento de informações para painéis

Os metadados de áudio podem ser inseridos em painéis analíticos, como VSaaS (videomonitoramento como serviço), IoT (internet das coisas) ou plataformas de inteligência empresarial, que reúnem e apresentam os metadados visualmente. Esses painéis analisam tendências históricas e em tempo real para gerar uma visão geral instantânea e visões práticas. As análises estatísticas baseadas no fluxo de clientes ou na experiência dos clientes possibilitam uma tomada de decisão orientada por dados, o que aprimora as operações.

Com os painéis, você verá os resultados sem precisar ouvir o áudio real ou recuperar o som original. Em vez disso, você poderá obter visões úteis, como contagem de eventos, e não terá dúvidas em relação à privacidade dos dados. Observe que pode haver diferenças nas restrições legais dependendo se o áudio é gravado ou apenas capturado.

## 4 Restrições legais

Muitas pessoas têm preocupações em relação ao uso de microfones em videomonitoramentos. Essas preocupações normalmente são vinculadas à gravação de fala simples juntamente com o material de vídeo. Mas, com os analíticos de áudio, os áudios geralmente não são gravados nem transmitidos. As leis que regulam o monitoramento variam de acordo com a região e o país, portanto, verifique o que é permitido antes de usar áudio no seu sistema de monitoramento.

A captura e a gravação de áudio podem ser proibidos ou exigir avaliação especial por diversas razões, por legislação nacional ou vários tipos de regras e regulamentos locais. Embora uma região ou ambiente possa permitir a captura de áudios, ela pode proibir a gravação de áudios. As empresas também podem proibir o monitoramento de áudio dentro das instalações.

## 5 Aviso de isenção de responsabilidade

Este documento e seu conteúdo são fornecidos como cortesia da Axis e todos os direitos ao documento ou quaisquer direitos de propriedade intelectual relacionados a ele (incluindo, mas não se limitando a marcas registradas, nomes comerciais, logotipos e marcas semelhantes neles) são protegidos por lei e todos os direitos, títulos e/ou interesse no e para o documento ou quaisquer direitos de propriedade intelectual relacionados a ele são e devem permanecer propriedade da Axis Communications AB.

Informamos que este documento é fornecido como se encontra ("as is"), sem garantia de qualquer tipo apenas para fins informativos. As informações fornecidas neste documento não constituem, nem pretendem constituir, aconselhamento jurídico. Este documento não se destina a, e nem deve, criar qualquer obrigação legal para a Axis Communications AB e/ou qualquer uma de suas afiliadas. As obrigações da Axis Communications AB e/ou de qualquer uma de suas afiliadas em relação aos produtos Axis estão sujeitas exclusivamente aos termos e condições do contrato entre a Axis e a entidade que adquiriu tais produtos diretamente com a Axis.

PARA EVITAR DÚVIDAS, TODOS OS RISCOS RELACIONADOS AO USO, RESULTADOS E DESEMPENHO DESTES DOCUMENTOS SÃO ASSUMIDOS PELO USUÁRIO DO DOCUMENTO, E A AXIS SE ISENTA E EXCLUI, NO LIMITE MÁXIMO PERMITIDO POR LEI, TODAS AS GARANTIAS, SEJAM ELAS PREVISTAS POR LEI, EXPLÍCITAS OU IMPLÍCITAS, INCLUINDO, SEM LIMITAÇÕES, QUALQUER GARANTIA IMPLÍCITA DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UM DETERMINADO FIM, TÍTULO E NÃO VIOLAÇÃO E RESPONSABILIDADE PELO PRODUTO, OU GARANTIAS DECORRENTES DE QUALQUER PROPOSTA, ESPECIFICAÇÃO OU AMOSTRA COM RELAÇÃO A ESTE DOCUMENTO.



# Sobre a Axis Communications

A Axis torna possível um mundo mais inteligente e seguro criando soluções para melhorar a segurança e o desempenho dos negócios. Como empresa de tecnologia de rede e líder do setor, a Axis oferece soluções em videomonitoramento, controle de acesso, intercomunicação e áudio. Nossas soluções são aprimoradas por aplicativos de análise inteligentes e apoiados por treinamento de alta qualidade.

A Axis tem cerca de 4.000 funcionários dedicados em mais de 50 países e colabora com parceiros de tecnologia e integração de sistemas em todo o mundo para fornecer soluções aos clientes. A Axis foi fundada em 1984 e tem sede em Lund, Suécia