AWS SUMMIT JAPAN 2024 - 製造ブースデモ展示

生成 AI によるカメラ映像からの危険判別

製造業における課題

生産部門 (内部)

自社従業員の安全確保

- 安全確保は最優先事項
- カメラで監視

人手不足の解消

- 監視カメラ映像を確認する労力がカメラ台数を制約
- 従来の機械学習では、検知できる事象が限定的

企画開発部門 (外部)

モノからコトへ

- 顧客価値を提供する最小構成(MVP)をすぐに市場に投入したい

映像・画像を扱う課題

- 映像転送、エッジやクラウドの配置
- 複数の機械学習の組み合わせ

AI/MLモデル開発の負担解消

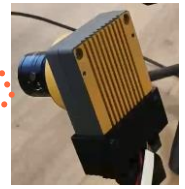
- 従来の機械学習では、転倒など、事象ごとに専用のモデル開発が必要

従業員の安全を守る Generative AI Camera

生成 AI を活用した リアルタイムのカメラ映像からの情報抽出とシーン解説

製造現場・工場など

現場の様子



カメラで
映像を撮影

映像をリアルタイムに
AWS に転送、
生成 AI や機械学習で解析

Kinesis Video Streams

Bedrock (Claude 3)

Rekognition

Web アプリ

ライブ動画

Caution

建設作業員が建設中の家屋内で梯子から落ちようとしていて、右手を伸ばしています。危険な事故が起きていると思われます。...

状況の説明

ユースケース 1 : 危険への即応

監視カメラ映像からの
事故や危険なイベントの抽出と通知

ユースケース 2 ：危険の予防

現場作業者の安全装備(ヘルメットやベスト等)の
カメラ映像によるチェック

ユースケース１：監視カメラによる遠隔監視

施設・ビルの警備

監視カメラを監視し、
脅威や未確認の出来事に
遠隔で対応する必要性

課題

- 複数のカメラを同時に監視する必要がある
- 手間のかかるビデオ解析が必要
- 低レベルのイベント通知が多すぎる

解決策

- エッジコンピューティング
  - エッジの機械学習モデルによるシンプルな検出
- 生成AI
  - 高レベルのシーン説明を提供
- クラウドへのビデオストリーミング
  - カメラからの複数のビデオストリームを遠隔監視

AWS IoT Greengrass

Amazon Bedrock

Amazon Kinesis Video Streams

ユースケース 2 - 作業現場での PPE 着用の確認

工場における安全管理

現場全体における
PPE 着用を徹底する必要性

※ PPE = 個人用保護具
(Protective Protection Equipments)

課題

- 様々な種類の PPE への対応が困難
- MLモデルの学習が困難
- コンプライアンスの報告が手作業または煩雑

解決策

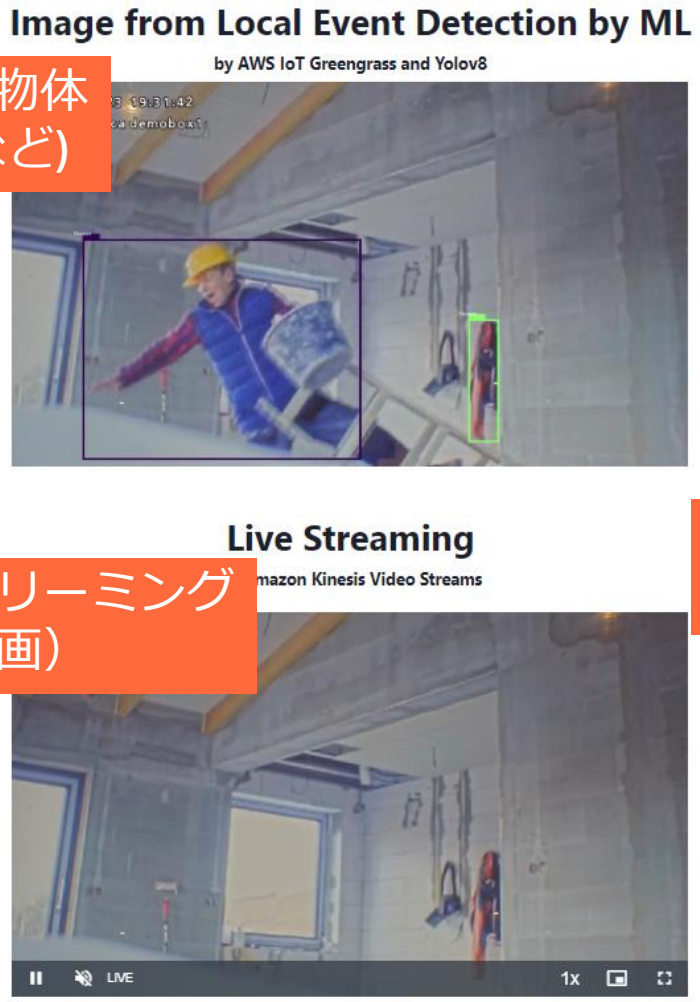
- **生成 AI（Image2Text)**
  - 自然言語での問合せで画像の内容を分析可能に
- クラウドへのビデオストリーミング
  - ビデオをクラウドに送信しリモートで確認

AWS IoT Greengrass
Amazon Rekognition
Amazon Bedrock
Amazon Kinesis Video Streams

デモアプリケーション画面の様子

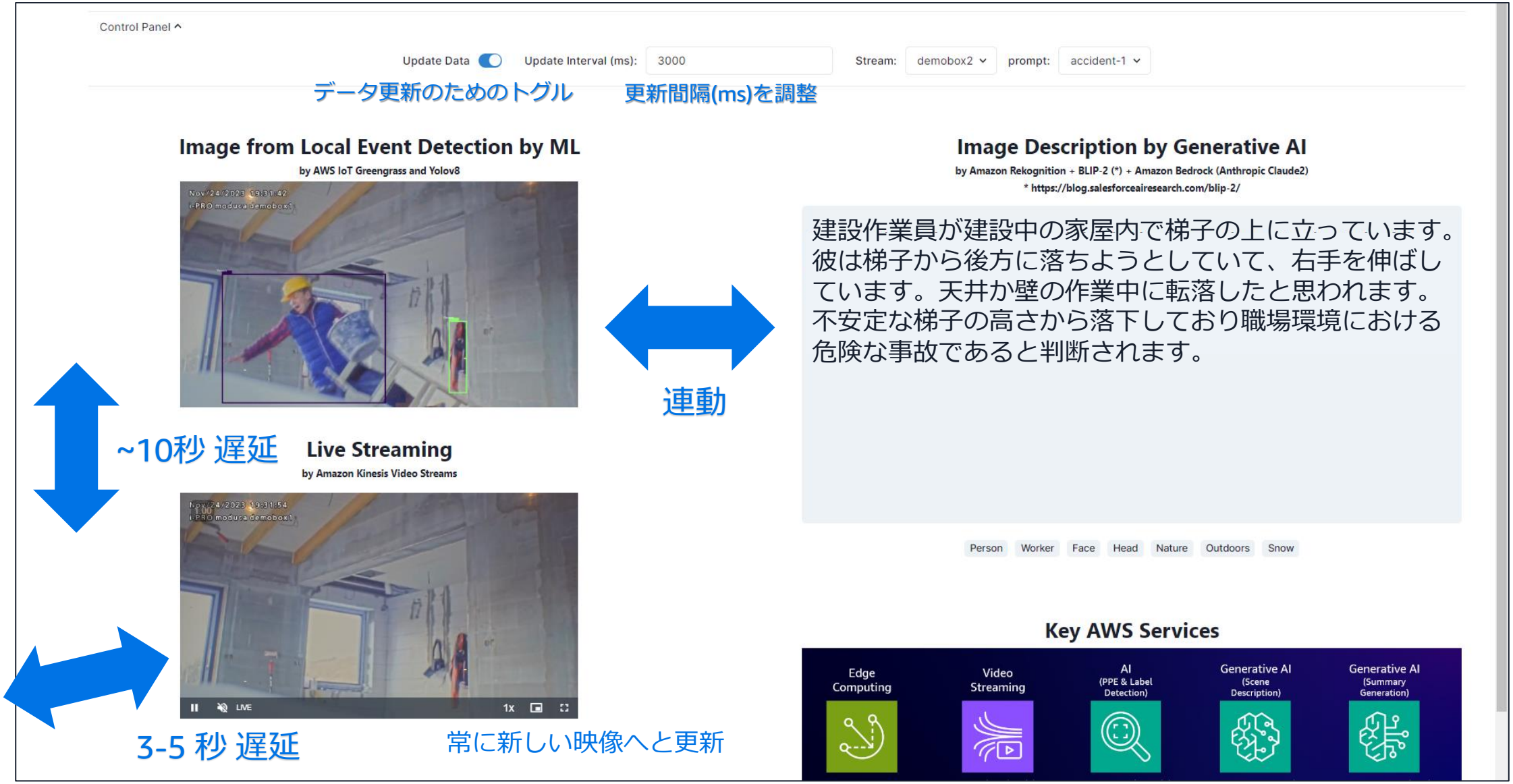
検出された物体
(人、装備など)

ライブストリーミング
（動画）

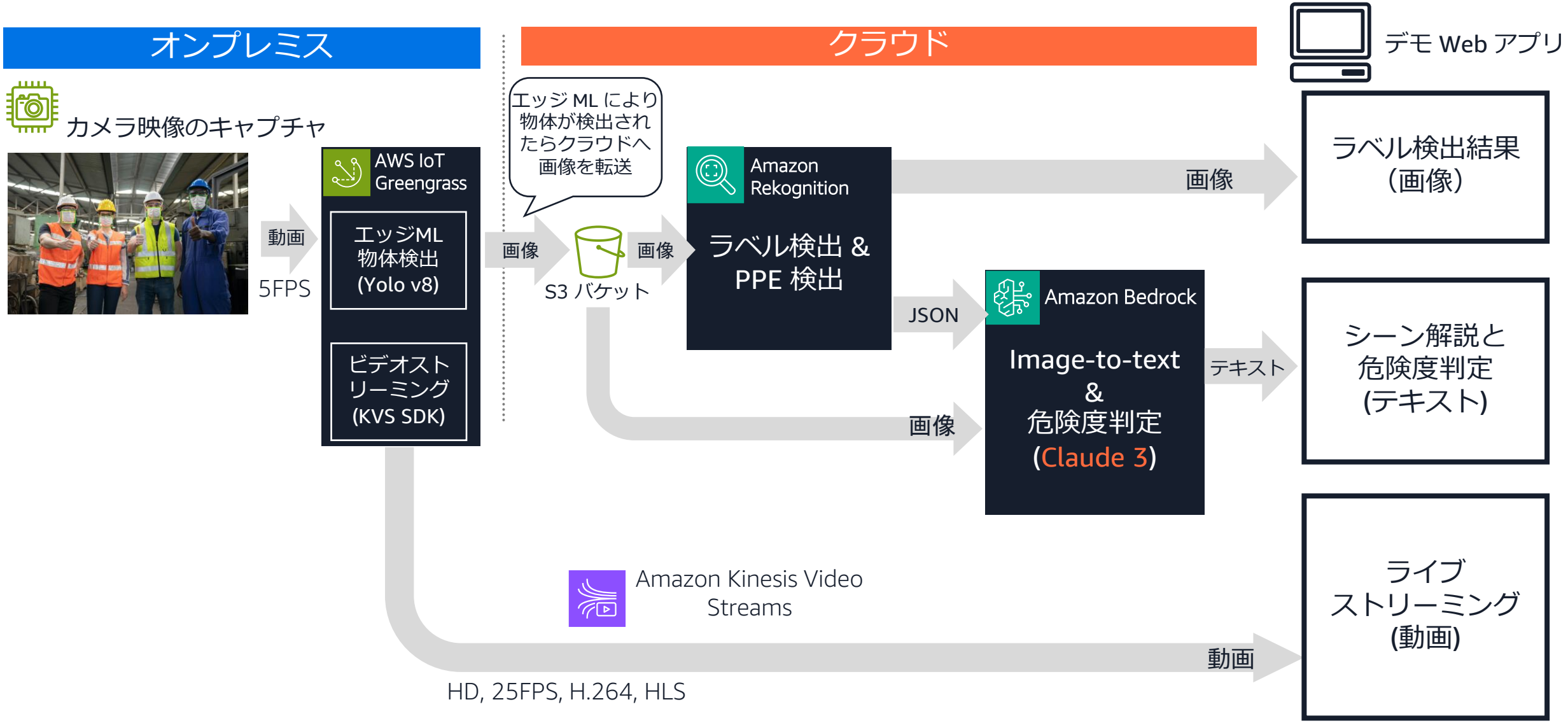


生成系AIによる
状況テキスト化

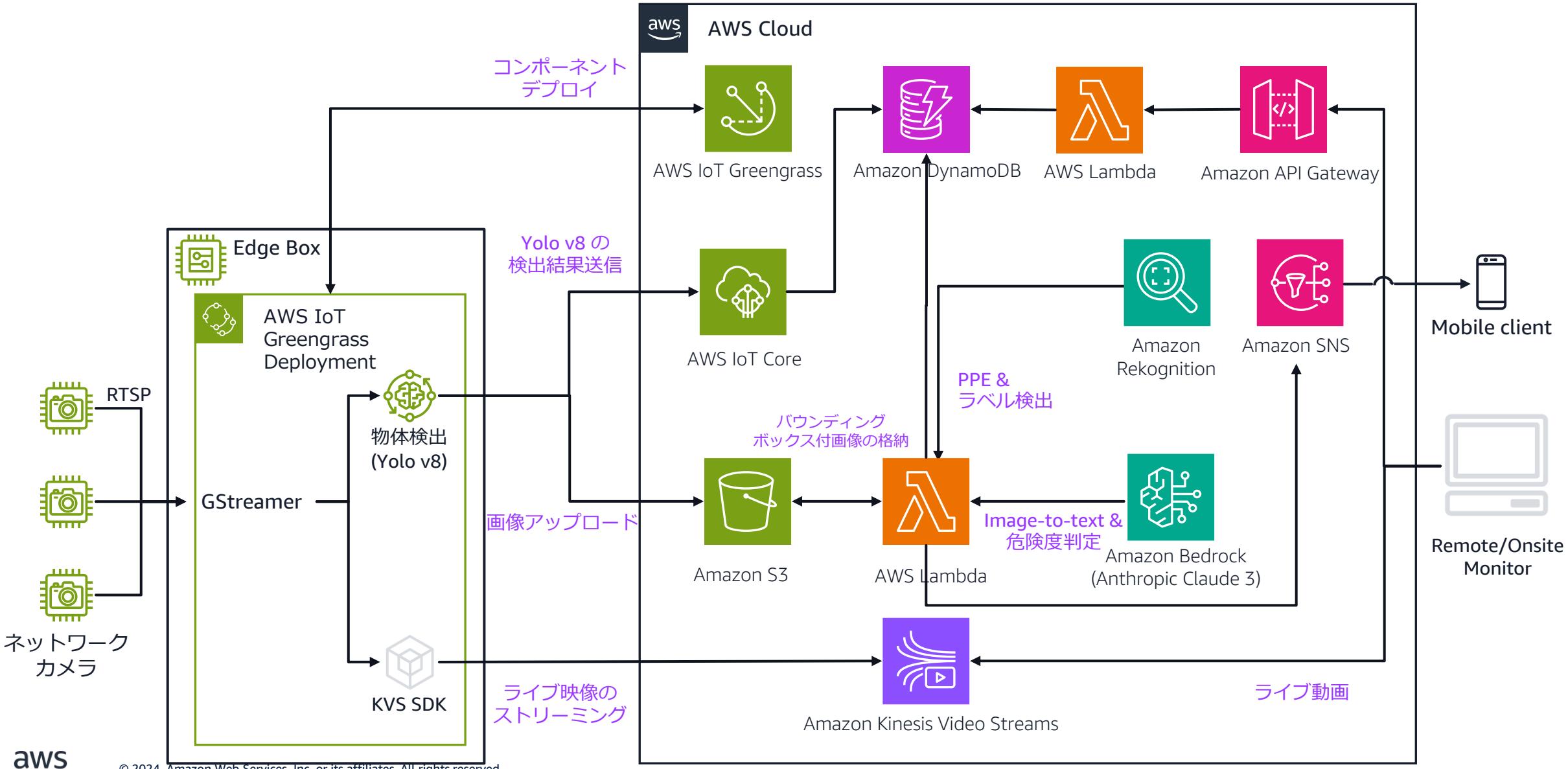
デモアプリケーション画面（遅延）



Smart Products Demo – 処理フロー概要



Smart Products Demo – アーキテクチャ詳細

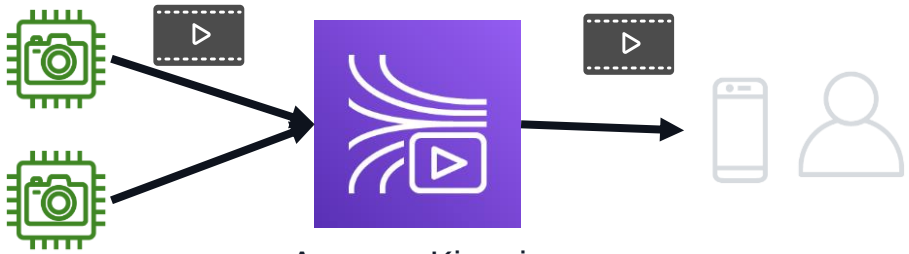


コスト試算

コスト試算

映像転送

Amazon Kinesis Video Streams



- SD (1280x720) 1.5Mbps/sec動画をクラウドに常時録画。
- 録画した映像のうち1日あたり1時間分をHLSで再生。
- 録画の保存期間は2週間とする。

カメラー台・一月あたり

動画の取り込み	動画の保存	動画の再生:
$5.2	$6.0	$2.2

13.4$ /camera

画像認識

Amazon Bedrock - Anthropic Claude 3

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku
ユースケース	最も知性が高く、最高のパフォーマンス	知能、スピード、コストのバランスがとれたモデル	最も低コストで最速のパフォーマンス
コスト* (* 1000トークンあたり)	Input: $0.015 Output: $0.075	$0.003 $0.015	$0.00025 $0.00125

今回のデモで使用

画像を入力した場合のトークン換算

tokens = (width px * height px)/750

Amazon Rekognition ※ オレゴンリージョン、2024年6月時点

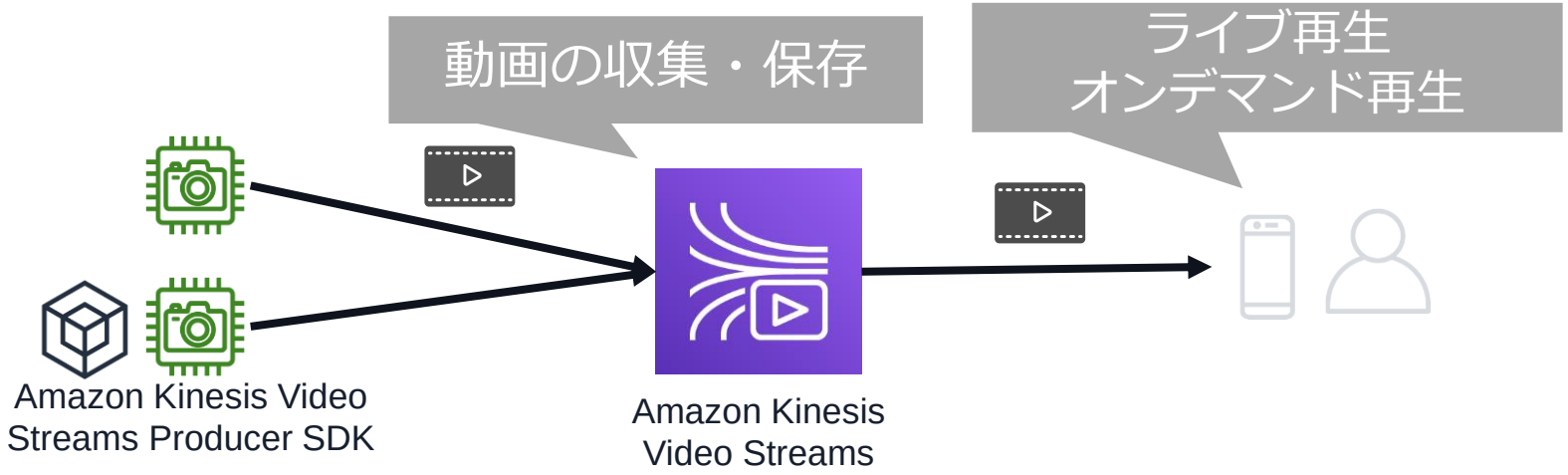
ラベル検出・PPE検出ともに
最初の100万枚まで $ 0.001 /枚

1回の呼び出しあたり

約 **$0.012** / image

24Hクラウド録画のコスト試算例: Amazon Kinesis Video Streams

Item	Value
接続台数	10,000台
稼働率	100%
ビットレート	0.75Mbps
データ保存	1ヶ月
月額 total	$98,070
月額/Device	$9.81



動画の取り込み: $26,657

動画の保存: $60,750

動画の再生: $10,663

完全に従量で計算できる
(台数xビットレートx利用時間)

- 伝送方式は Stream
- SD (720x480) 0.75Mbps/sec程度のカメラ動画をクラウドに常時録画。
- 録画した映像のうち1日あたり1時間分をHLSで再生。
- 録画の保存期間は1ヶ月とする。

本資料はコスト試算例であり見積もりではありません。2023年3月17日時点のAWSサービス内容および東京リージョンの価格を元にしています。
必要に応じて項目などを調整してご活用ください。最新の情報はAWS公式ウェブサイトにてご確認ください。
料金表: https://aws.amazon.com/jp/kinesis/video-streams/pricing/
Pricing Calculator: https://calculator.aws/#/addService/KinesisVideoStreams

デモ – 画像解析 & テキスト生成 コスト

Amazon Bedrock - Anthropic Claude 3

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku
ユースケース	最も知性が高く、最高のパフォーマンス	知能、スピード、コストのバランスがとれたモデル	最も低コストで最速のパフォーマンス
コスト* (* 1000トークンあたり)	Input: $0.015 Output: $0.075	$0.003 $0.015	$0.00025 $0.00125

画像を入力した場合のトークン換算

tokens = (width px * height px)/750

Amazon Rekognition

※ オレゴンリージョン、2024年6月時点

ラベル検出・PPE検出ともに
最初の100万枚まで $ 0.001 /枚

今回のデモの場合（1回の呼び出し）

- 画像解像度：1280 x 720
- Claude3 Sonnet
- プロンプト (Appendix 参照)
- Amazon Rekognition ラベル検出・PPE検出

- Amazon Bedrock : 約 $0.01
  - 入力トークン数：約2500～3000
    (画像 1228 トークン)
  - 出力トークン数：約100
- Amazon Rekognition ：$ 0.002

約 **$0.012** / request

Key AWS Services

Key AWS Services

AWS IoT
Greengrass

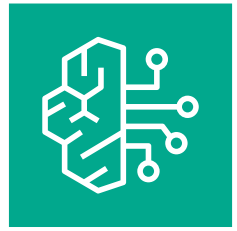
エッジ
コンピューティング

Amazon Kinesis
Video Streams

ビデオ
ストリーミング

Amazon Rekognition

画像認識 AI
(PPE & ラベル 検出)



Amazon Bedrock
(Anthropic Claude 3)

生成 AI
(シーン解説 & 要約生成)

AWS IoT Greengrass

AWS IoT Greengrass は、AWS をお使いのデバイス上に拡張することによって、クラウドの利点を活かしながら、データのローカルでの処理を可能にします。

Amazon Kinesis Video Streams

クラウドへ動画をリアルタイムに簡単かつ安全にストリーミング

Streams

Streams

カメラデバイスからのデータ取り込み

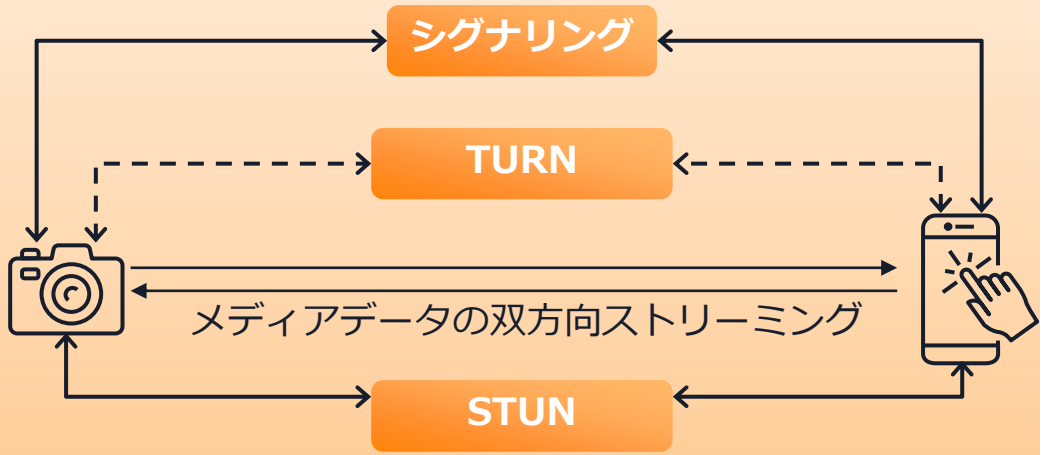
メディアを取り込み、保存、消費、
タイムインデックス付きメディアデータを再生
AI/ML サービスとの統合

WebRTC

WebRTC

低遅延で双方向の
メディアストリーミング

マネージドのシグナリング、STUN、TURN サーバ



Amazon Rekognition

深層学習をベースにした画像・動画認識 AI サービス

物体・シーン検出

顔検出・分析

顔の比較

有名人認識

コンテンツの
モデレーション

保護具検知

テキストの検出

人物の動線追跡

動画のシーン分析

顔検索

動画ストリーミングの分析

カスタムラベル

Face Liveness

Amazon Bedrock

2024年6月時点

幅広い基盤モデルの選択肢をご提供

JURASSIC

AI21 labs

テキスト

Jurassic-2 Ultra
Jurassic-2 Mid

CLAUDE

ANTHROPIC

テキスト & ビジョン

Claude 3.5 Sonnet
Claude 3 Opus
Claude 3 Sonnet
Claude 3 Haiku

テキスト

Claude 2.1
Claude 2.0
Claude Instant

MISTRAL & MIXTRAL

MISTRAL AI_

テキスト

Mistral Large
Mistral Small
Mistral 7B
Mixtral 8X7B

COMMAND & EMBED

co:here

テキスト

Command R+
Command R
Command
Command Light

埋め込み

Embed - Multiligual
Embed - English

LLAMA

Meta AI

テキスト

Llama 3 70B
Llama 3 8B
Llama 2 70B
Llama 2 13B

SDXL

stability.ai

画像

Stable Diffusion XL 1.0

AMAZON TITAN

amazon

テキスト

Titan Text Premier
Titan Text Express
Titan Text Lite

画像

Titan Image Generator

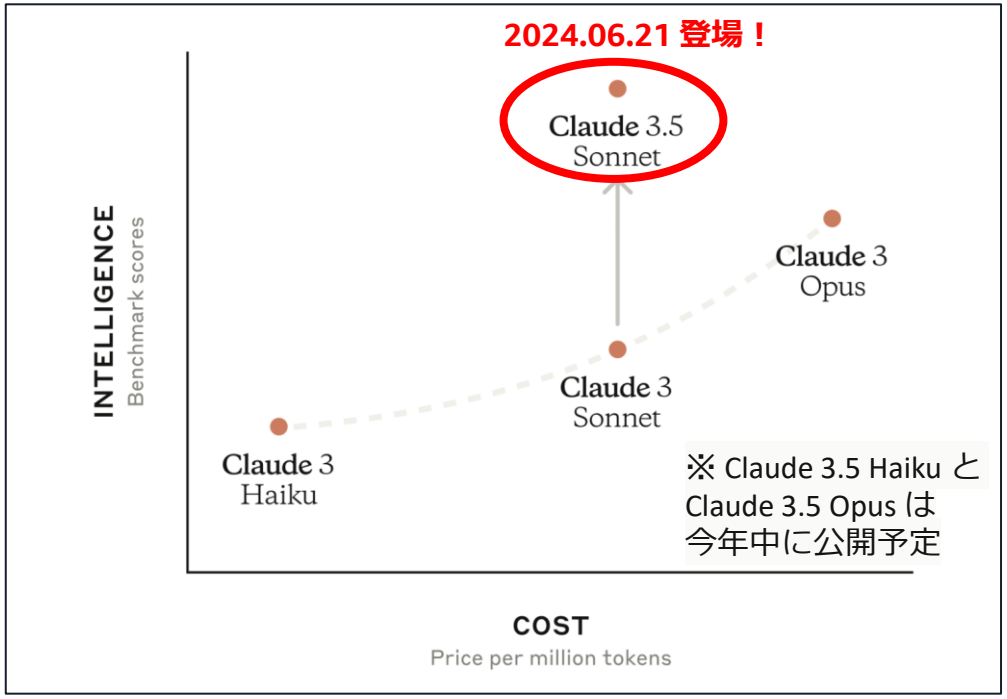
埋め込み

Titan Multimodal Embeddings
Titan Text Embeddings V2
Titan Text Embeddings

aws

Amazon Bedrock における Claude 3.5/3 family

用途に合わせて 知能、スピード、コストの組み合わせを選択可能



	Claude 3.5 Sonnet	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku
ユースケース	最もインテリジェントなモデルで、トップクラスのパフォーマンスと向上したスピード	非常に複雑なタスクで優れたパフォーマンスを発揮	知能、スピード、コストのバランスがとれたモデル	最も低コストで最速のパフォーマンス
コンテキスト長	200k	200K	200K	200K
ビジョン対応	✓	✓	✓	✓

性能表

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro	Llama-400b (early snapshot)
Graduate level reasoning *GPQA, Diamond*	59.4%* 0-shot CoT	50.4% 0-shot CoT	53.6% 0-shot CoT	–	–
Undergraduate level knowledge *MMLU*	88.7%** 5-shot 88.3% 0-shot CoT	86.8% 5-shot 85.7% 0-shot CoT	– 88.7% 0-shot CoT	85.9% 5-shot –	86.1% 5-shot –
Code *HumanEval*	92.0% 0-shot	84.9% 0-shot	90.2% 0-shot	84.1% 0-shot	84.1% 0-shot
Multilingual math *MGSM*	91.6% 0-shot CoT	90.7% 0-shot CoT	90.5% 0-shot CoT	87.5% 8-shot	–
Reasoning over text *DROP, F1 score*	87.1 3-shot	83.1 3-shot	83.4 3-shot	74.9 Variable shots	83.5 3-shot Pre-trained model
Mixed evaluations *BIG-Bench-Hard*	93.1% 3-shot CoT	86.8% 3-shot CoT	–	89.2% 3-shot CoT	85.3% 3-shot CoT Pre-trained model
Math problem-solving *MATH*	71.1% 0-shot CoT	60.1% 0-shot CoT	76.6% 0-shot CoT	67.7%	57.8%
Grade school math *GSM8K*	96.4% 0-shot CoT	95.0% 0-shot CoT			

マルチモーダル（画像入力）に対応

Thank you!

Prompt Template (Usecase: Accident)

<instruction>
You are an AI assistant tasked with analyzing images and determining if any accident is occuring in the scene. You will be provided with an image, the results of an image recognition model, and the results of a personal protective equipment (PPE) detection model.

Your task is to:
1. Answer the following question: "Tell us what the situation is like with this image in detail. Is there any trouble going on? Generate captions in more than 3 sentences." (image_caption)
2. Determine if there is an on-going occuring trouble or dangerous situation(classification), and output either 0 (No) or 1(Yes).

Your output must be formatted as a JSON object with image_caption and classification keys.
{{
  "image_caption": "<caption here>",
   "classification": <0 or 1>
}}
Please provide your analysis based on the given inputs.
</instruction>

<rekognition_label>{rekognition_label}</rekognition_label>

<rekognition_ppe>{rekognition_ppe}</rekognition_ppe>

<reference>When answering the question related to the position of the image, you can use tha fact that a value of 'Left' closer to 0.0 indicates the left side of the image, closer to 0.50 the middle, and closer to 1.0 the right side. And you can find the numbers of people in "Number of Persons' of rekognition_ppe.</reference>

<outputRule>The final output should be by JSON and any other characters except JSON object is prohibited to output. </outputRule>

<outputLanguage>In Japanese.</outputLanguage>