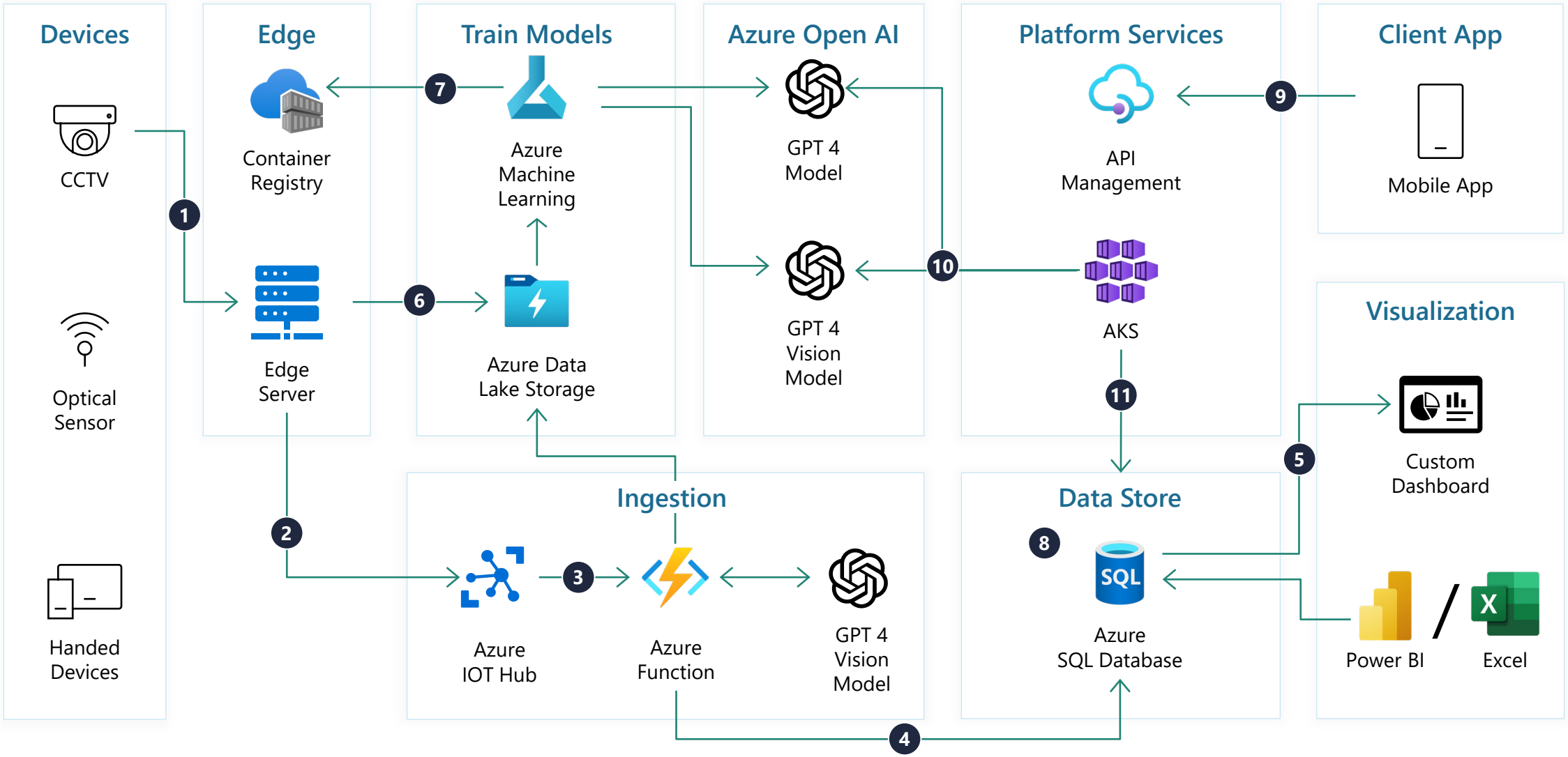
Multi-Modal Image Analysis

Multi-Modal Image Analysis architecture



Multi-Modal Image Analysis architecture

1. **Video Acquisition and Initial Processing:** A Closed-Circuit Television (CCTV) camera captures video in real-time using the Real-Time Streaming Protocol (RTSP). The live stream is fed into a containerized application deployed on an Edge Server. This application processes the video stream, extracting individual frames at key intervals. These key frames are then saved to an Azure Data Lake Storage account, which acts as a repository for the raw data.

2. **Metadata Forwarding:** Along with the image data, metadata is generated which might include timestamps, frame identifiers, or other contextual information. This metadata is sent to the Azure IoT Hub, which serves as a central point for data ingestion from various IoT devices and sensors.

3. **Image Analysis with AI:** The Azure Function is triggered with the arrival of new data in the IoT Hub. It invokes OpenAI's vision models, which are pretrained generative models capable of analyzing images for anomalies or specific features. These AI models process the frames to detect any irregularities or incidents, such as the anomaly detected at the East gate in Lisa's use case.

4. **Data Persistence:** The results of the analysis, which could include detections, classifications, and other relevant AI insights, are stored in a database. This database serves as a structured storage solution, enabling efficient querying and data management.

5. **Data Consumption and Visualization:** The stored data can be accessed by various visualization tools to provide a comprehensive operational dashboard. A custom application is developed for Lisa, equipped with a user interface that includes command controls and alert mechanisms. Integration with services like Microsoft Power BI allows for advanced data visualization and reporting. Additionally, the data in the database can be exported and analyzed using tools like Microsoft Excel, offering further flexibility for data analysis.

Multi-Modal Image Analysis architecture

6 **Data Aggregation:** The Azure Data Lake Storage acts as a comprehensive repository, housing all videos and sensor data from the previous month. This includes raw footage from CCTV cameras, which captures the daily activities within the monitored premises.

7 **Model Training and Enhancement:** Leveraging both Azure OpenAI models and custom machine learning models, a new set of analytics tools is created. These new models are trained using the aggregated data, focusing on identifying patterns of activity.

The models are likely trained to recognize varying levels of human activity and correlate it with staffing requirements.

8 **Data Storage and Accessibility:** Insights from the model, such as identified patterns and detected anomalies, are stored back into the data layer, likely within a structured database for efficient querying.

This storage layer serves as a historical record of insights and analytics derived from the AI models.

Monitoring tools can utilize this data to generate alerts or reports that highlight critical insights, such as the need for staffing adjustments during peak activity hours.

Multi-Modal Image Analysis architecture

9. Workers use a mobile application to capture a photograph of an empty shipping vessel. This application seamlessly interfaces with Azure Kubernetes Service (AKS) and Azure API Management. This integration ensures efficient processing and secure management of the requests initiated by the mobile application.

10. Upon receiving the photograph, the integrated services leverage Azure OpenAI Vision models to analyze the image. These advanced AI models are designed to assess the empty space within the shipping vessel and generate optimal placement suggestions. The objective is to maximize the utilization of the available space. This process involves intricate analysis of the photograph to derive a detailed and efficient layout for cargo arrangement.

11. In addition to the AI-based analysis, the services also access a SQL database for necessary reference data. This database contains historical patterns, physical dimensions of the shipping vessel, weight limits, or other relevant information that aids in refining the placement suggestions. The integration with the database ensures that the recommendations are not only based on the current image analysis but are also informed by past data and constraints, leading to more accurate and practical suggestions for cargo arrangement.

Multi-Modal Image Analysis limitations

Spatial reasoning ⚠

The model struggles with tasks requiring precise spatial localization, such as identifying chess positions.

Metadata and resizing

The model doesn't process original file names or metadata, and images are resized before analysis.

Image shape ⚠

The model struggles with panoramic and fisheye images.

Counting ⚠

May give approximate counts for objects in images.

Rotation

The model may misinterpret rotated/upside-down text or images.