Proprietary video from sources like VMS, Transit, Body Cameras, Covert Cameras, Consumer DVRs, Doorbell Cams, and more are flooding digital evidence / investigation processes. AI analysis tools and Redaction Engines do not operate on digital media files and associated metadata created in proprietary formats. The workflow to process proprietary video into standardized media and metadata is time consuming and very manual in nature. Uploading large media files to the Cloud for processing is format dependent, bandwidth intensive, expensive, and slow.

In many workflow evidence management solutions, the process is “upside down” in that media has to be uploaded to a cloud service in bulk volume in order to be able to process, convert, or open it. If you have 30 days of video from 4 cameras that can be a very large file-based upload. By processing the video immediately on premise to open and review what is evidentiary, running AI tools on the content such as BriefCam, finding what's most important for the investigation, and then uploading the results is 90% or more faster and less expensive.

Recommendations for Digital Investigations In Modern Evidence Systems

1. **Improve Data Quality and Consistency**

- Incorporate Automatic Video Conversion and Data Cleansing on-premise to improve data integrity and accuracy.
- Data Cleansing - Standardize metadata to ensure that all data points, regardless of their source, follow a consistent structure, terminology, and format.
- AI models require clean, standardized, and well-organized media and data for accurate predictions and analyses. Standardization minimizes errors, duplicates, and inconsistencies, leading to higher-quality results less prone to generating hallucinations.

2. **Simplified Data Preparation and Preprocessing**

- AI algorithms use cases can require a significant amount of time in data “wrangling” (cleaning, organizing, and preprocessing video and data).
- With standardized video and metadata, attributes like file names, tags, creation dates, and relationships are uniformly defined, streamlining tasks like feature extraction, labeling, and data transformation.
- Centralized assets reduce the need to merge datasets from disparate locations, saving time and resources during the data preparation phase.

3. **Scalability and Automation**

- With standardized media and metadata, AI systems are more accurate, easier for the end user, and enhance the ability to use AI for Compliant Legal Processes.
- Centralized access powers models to scale efficiently across large diverse datasets from multiple vendors and sources.

4. **Better Integration with AI Tools and Frameworks**

- Standardized media and metadata make integration with AI tools such as BriefCam, AWS, Microsoft Minted, Dell, Centific, and others.
- Centralized data repositories support Data Lake architectures and feature storage repositories, both critical for modern AI workflows.
- Use APIs to query standardized metadata for specific data points, enabling dynamic model retraining or real-time analytics.

5. **Enhanced Compliance and Ethics in AI**

- Standardized metadata helps ensure compliance with data governance policies and regulations, such as GDPR, CCPA, or HIPAA.
- Metadata tracks the source and permissions of assets, ensuring AI models are trained on ethical and legally compliant datasets.
- Retention Management ensures assets and metadata are preserved according to compliance standards and organizational needs.
- Audit trails maintained through metadata can document how data was used, aiding in explainability and transparency of AI systems.

Summary Statement

Software that cleanses media and metadata are the precursor to support the ingestion, metadata enrichment, validation, and delivery of digital evidence from various media and systems sources to designated backend platforms, AI Services, and Archiving. Without a strategy to cleanse media and data and store it in an accessible format whether on prem, cloud, or hybrid environments will ultimately drive the success or failure of most AI strategies.