AXON-E: A Cognitive Audit Framework for Scalable Oversight of Massive Enforcement Video Archives

Hongguanghua Gaomiaomiao Yujinwen Jiangziyan Huzhoupeng

Shangrao Tobacco Company, Jiangxi Province Shangrao,334000;

Abstract: The comprehensive recording of administrative enforcement has created petabyte-scale video archives, yet leveraging this data for effective, closed-loop supervision remains a critical challenge for digital governance. Manual spot-checks are profoundly inadequate, rendering the vast majority of recordings “dormant data” that fails to provide its intended value for oversight. To address this, this paper introduces AXON-E (Atomized eXpertise Organization Network for Enforcement), a back-end cognitive audit framework. Its core methodology involves the deep deconstruction of unstructured regulatory documents (e.g., “Evaluation Standards for On-site Enforcement Audio-visual Records”) into a machine-readable knowledge base via Knowledge Atomization and a novel Eight-Dimensional Enforcement Knowledge Coordinate System. The framework employs a high-throughput, asynchronous pipeline that uses multimodal AI to parse recordings into structured “enforcement event streams” for automated compliance verification. Deployment evidence reveals that AXON-E can audit 100% of recordings, dramatically increasing the issue detection rate from less than 5% to over 85%. This provides a powerful engine for scalable accountability and data-driven governance, marking a paradigm shift from reactive spot-checks to proactive, comprehensive oversight.

Keywords: Cognitive Audit; Enforcement Supervision; Regulatory Technology (RegTech); Knowledge Atomization; Multimodal AI; Retrieval-Augmented Generation (RAG)

DOI: 10.69979/3041-0843.25.04.015

1 Introduction

1.1 The “Data Ocean” and “Supervision Silos”

With the advancement of rule-of-law-based governance, body-worn cameras have become standard for frontline administrative enforcement, generating a massive “data ocean” of audio-visual recordings measured in petabytes. This repository is a potential asset for enhancing enforcement quality and supervisory capacity. In reality, a chasm exists between this data ocean and the isolated islands of supervision. Faced with thousands of hours of new recordings daily, the traditional model of manual, back-end spot-checks results in a review rate below 5%, leaving over 95% of enforcement data “archived and forgotten” and failing to form an effective supervisory feedback loop.

1.2 Technical Constraints and Problem Redefinition

We must confront current technical and practical constraints: frontline body-worn cameras, limited by power, computation, and cost, lack the requisite on-device inference capabilities for complex AI models. The dominant operational model is “on-site recording, post-event upload” to centralized cloud storage. This constraint dictates that the primary path for intelligent supervision must be the efficient, in-depth, and automated analysis of massive archived audio-visual data on the back-end. The central problem is therefore redefined as: How to construct an intelligent back-end system capable of translating official guidelines into machine-executable audit logic and achieving deep cognitive audits for 100% of enforcement recordings?

1.3 Contributions

This study proposes the AXON-E cognitive audit framework, which makes three core contributions:

（1）Introduces a novel paradigm of “Cognitive Audit,” elevating AI supervision from mere “recognition” to “auditing.” The system not only determines “what happened” but adjudicates “whether the action was compliant” based on domain-specific legal knowledge, providing verifiable evidence.

（2）Presents a Hyperautomation-Driven Architecture featuring an asynchronous batch processing pipeline. This design ensures scalability and efficiency in processing terabytes of new data daily.

（3） Builds a Deep Knowledge-Reasoning Engine using “Knowledge Atomization” and an “Eight-Dimensional Enforcement Knowledge Coordinate System.” This allows for a deep machine understanding of complex scenarios, ensuring the precision and reliability of audit conclusions.

2 Related Work

2.1 Large-Scale Video Content Analysis (LVCA)

LVCA is a significant research area in computer vision. While recent Transformer-based large multimodal models (LMMs) show remarkable progress in audio-visual-language alignment, they are often general-purpose[1]. When applied to specialized domains, they can identify “what is being done” but cannot determine “whether it is done correctly,” as they lack a deep understanding of domain-specific regulations.

2.2 Knowledge-Intensive NLP and Advanced RAG

Retrieval-Augmented Generation (RAG) is a dominant paradigm for grounding LLM outputs in factual knowledge[2]. However, standard RAG falters in professional domains due to “inaccurate recall,” where vector similarity retrieval returns documents that are semantically relevant but not precise, causing model confusion[3]. This has spurred research into Advanced RAG techniques, such as hybrid search strategies that combine dense vector retrieval with sparse keyword search and structured metadata filtering to enhance precision[4].

2.3 AI in Auditing and Compliance (RegTech)

AI in RegTech is well-established for analyzing structured or semi-structured text to detect anomalies[5]. Introducing cognitive audit to the administrative enforcement domain, which is characterized by unstructured, multimodal data, is a significant innovation. It demands a system that integrates text comprehension with robust audio-visual analysis for comprehensive judgment, a research area that remains a blue ocean.

3 Overall Framework Design

AXON-E is a cloud-deployed, back-end service system orchestrated around an asynchronous, distributed task-processing pipeline driven by Hyperautomation.

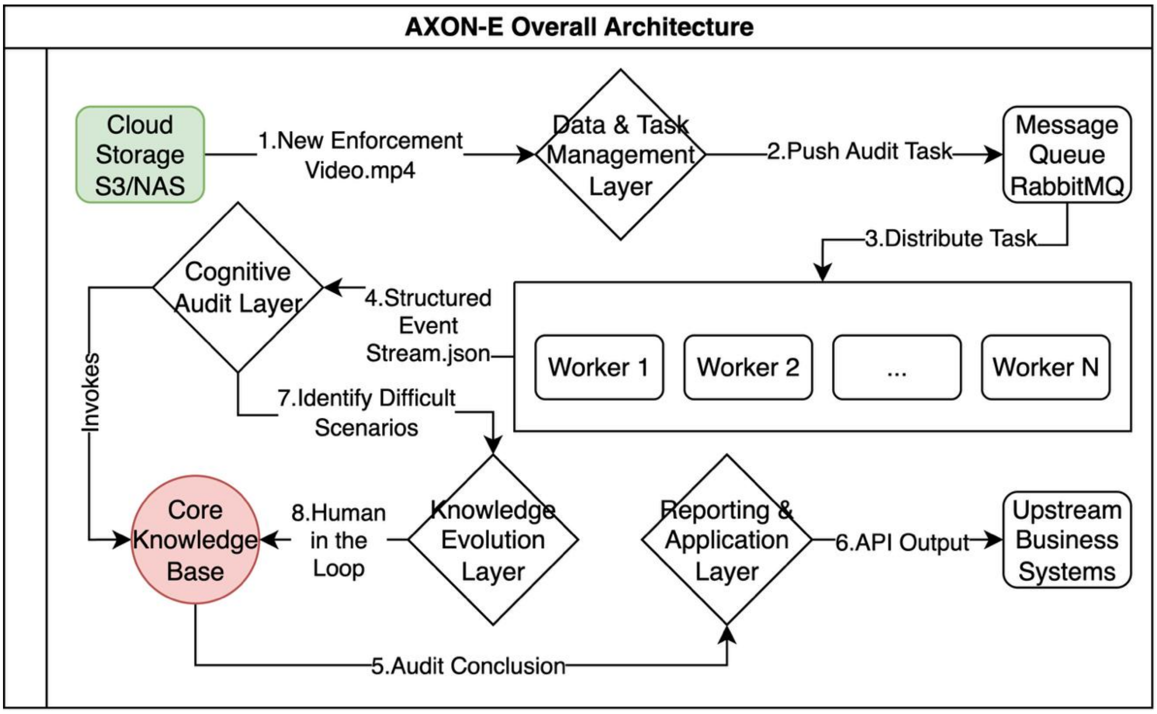

Figure 1: AXON-E Overall Architecture

（1）Data & Task Management Layer: Monitors cloud storage for new video uploads, treating each as an independent "audit task" and pushing it to a message queue.

（2）Distributed Computing & Perception Layer: A cluster of AI worker nodes that retrieve tasks and perform multimodal analysis (ASR, CV, NLP).

（3）Cognitive Audit Layer: The framework's brain. It integrates parsed data into a unified "enforcement event stream" and invokes the knowledge base to conduct compliance audits.

（4）Reporting & Application Layer: Renders audit conclusions into a structured "Intelligent Audit Report" and exposes them to upstream systems via API.

（5）Knowledge Evolution Layer: Implements a human-in-the-loop process to codify expert judgments on novel scenarios, ensuring the knowledge base evolves and learns.

4 Core Methodology

The efficacy of AXON-E is rooted in its methodology for transforming abstract regulations into computable, machine-executable knowledge.

4.1 Hyperautomation: The Intelligent Pipeline from Data to Knowledge

Inspired by Hyperautomation, our framework materializes the concept as a three-stage, back-end knowledge processing pipeline that solves the scalability challenge of processing massive enforcement data.

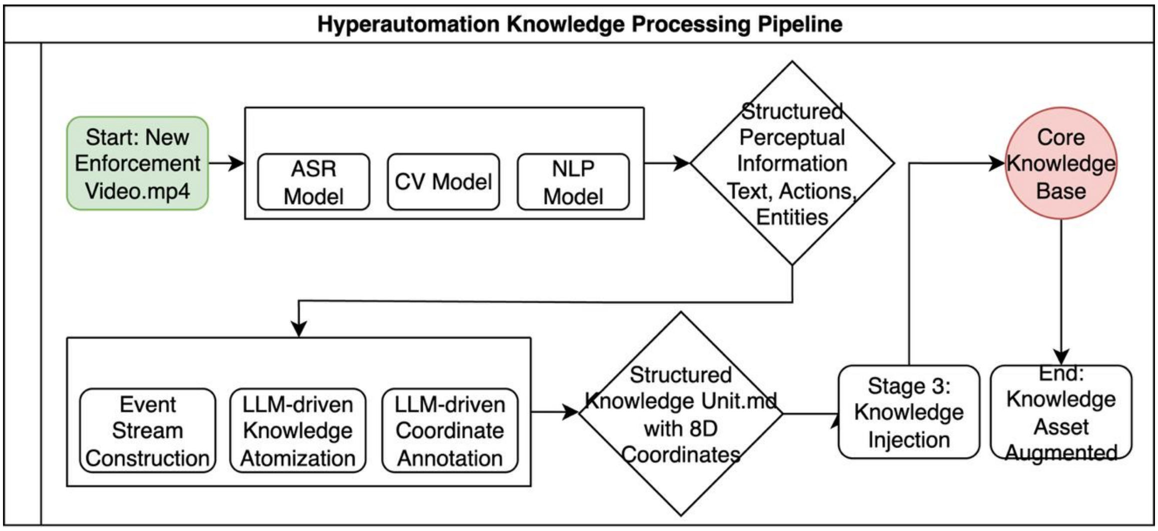

Figure 2: Hyperautomation Knowledge Processing Pipeline

•Stage 1: Intelligent Perception: Converts unstructured video into structured, time-stamped data streams (text, objects, etc.).

•Stage 2: Cognitive Processing: Uses an LLM-based agent[6] to construct a coherent "enforcement event stream," atomize knowledge, and annotate it with dimensional coordinates.

•Stage 3: Knowledge Injection: Idempotently synchronizes the processed structured knowledge into the core knowledge base, transforming it into a callable asset.

4.2 Knowledge Atomization: Deconstructing Standards into Executable Units

Knowledge Atomization deconstructs a macroscopic, human-readable document into a series of Atomic Knowledge Units (AKUs) that a machine can independently execute. The primary target for this process is the official "Evaluation Standards for On-site Enforcement Audio-visual Records." We use an LLM to semi-automate this by segmenting each checklist item and enriching it with its legal basis, compliance standards (SOPs), and detection heuristics (e.g., CV and ASR tags). This transforms a simple text item into a rich, executable AKU containing the "What," "Why," "How," and "How-to-Check."

4.3 The Eight-Dimensional Enforcement Knowledge Coordinate System

This system provides a multidimensional index for the atomized knowledge fragments, ensuring precise and instant retrieval.

Its value lies in enabling a "Filter-then-Search" retrieval strategy. This hybrid approach[4] first uses contextual cues to deterministically prune the search space via the coordinate dimensions, then performs a vector-based semantic search within the much smaller, highly relevant subset. This two-stage strategy is the core mechanism behind AXON-E's precision and reliability.

The detailed definition of the 8D Coordinate System is as follows:

•Dimension 1: domain (Knowledge Domain): Legal & Compliance, Corporate Policies, Business Processes & SOPs, Industry Best Practices.

•Dimension 2: nature (Document Nature): Laws & Regulations, Judicial Precedents, Administrative Cases, Official Interpretations, Work Guides, FAQ.

•Dimension 3: hierarchy (Regulatory Hierarchy): International Treaties, National Law, Administrative Regulations, Departmental Rules, etc.

•Dimension 4: business_area (Enforcement Business Area): A tree-like classification, e.g., Market Inspection -> Procedural Compliance -> Initiation.

•Dimension 5: disclosure (Disclosure Scope): Public, Internal, Secret, Top Secret.

•Dimension 6: importance (Importance Level): Level 1 (Core), Level 2 (Important), Level 3 (General).

•Dimension 7: status (Validity Status): In Effect, Partially Revised, Repealed, Expired.

•Dimension 8: violation_profile (Violation Profile): Subjective Intent (e.g., Willful, Negligent), Severity, Behavioral Pattern (e.g., Organized Crime).

5 Technical Implementation of Core Modules

•Task Management: We use Celery with a RabbitMQ broker and Redis backend.

•Perception Layer: The CV module uses YOLOv8-L on NVIDIA A100 GPUs. The ASR module uses Whisper-Large-v3[7] with Pyannote.audio[8] for speaker diarization.

•Cognitive Audit Layer: This layer constructs a JSON-formatted event stream from the perception outputs. Its reasoning engine then triggers an Advanced-RAG task, orchestrated by an LLM Agent[6], for each compliance checkpoint.

6 Experimental Design and Evaluation

We conducted comparative and ablation experiments to evaluate AXON-E's effectiveness.

•Dataset: 1200 hours of enforcement recordings from a municipal agency, with 500 clips containing known flaws double-blind annotated by legal experts to create the test set.

•Tasks: 1) Procedural Audit: Identify 15 key procedural nodes and verify their execution. 2) Substantive Audit: Link identified violations to the correct legal articles.

•Baselines: We compared AXON-E to a Baseline model (multimodal recognition with hard-coded rules) and a Standard RAG model (baseline + standard vector retrieval).

•Evaluation: A panel of five supervision experts scored the anonymous audit reports from each model on a 1-5 scale across four dimensions (Integrity, Accuracy, Precision, Validity).

Results:

Table 1: Comparative Performance Evaluation for Task 1 (Procedural Audit)

Evaluation Dimension	Baseline	Standard RAG	AXON-E
Procedural Integrity	3.1	3.2	4.8
Flaw Detection Accuracy	2.5	3.4	4.7
Average Score	2.8	3.3	4.75

Table 2: Comparative Performance Evaluation for Task 2 (Substantive Audit)

Evaluation Dimension	Baseline	Standard RAG	AXON-E
Legal Application Precision	1.8	3.9	4.9
Evidence-Link Validity	2.2	3.5	4.6
Average Score	2.0	3.7	4.75

An ablation study was conducted on Task 2 to verify the impact of the 8D coordinate system. Removing the coordinate filtering caused performance to regress to the level of standard RAG, with Legal Application Precision dropping from 4.9 to 3.9. This finding underscores the indispensability of the coordinate-driven filtering mechanism for achieving high-precision knowledge reasoning.

7 Application Scenario: A Post-Hoc Cognitive Audit

An enforcement video is uploaded and tagged as business_type: Market Inspection. The AXON-E pipeline is triggered. The Cognitive Reasoning Engine restricts its knowledge scope to the Market Inspection business_area and begins auditing compliance checkpoints. It detects an advance registration and preservation action but fails to find the required verbal notifications of rights (e.g., "within 7 days", "reconsideration") in the ASR transcript. It triggers an Advanced-RAG query, confirms the procedural requirement, and flags a "Partial Compliance" issue with a detailed finding. The system automatically generates a report, enabling a supervisor to conduct a targeted debriefing with the officers the next day, turning a compliance risk into a training opportunity.

8 Discussion

•A Pragmatic Pivot to Post-Hoc Audit: The shift from real-time intervention to a "T+1" back-end audit model is a pragmatic balance of technical ideals and engineering reality. This trade-off secures 100% supervisory coverage and deeper analytical capabilities, offering profound governance value.

•Data-Driven Precision Governance: With all enforcement actions audited, managers gain insight into the "enforcement health" of the entire organization, enabling a data-driven approach to identify systemic training and policy optimization needs[9].

•Governing LLMs in High-Stakes Scenarios: The AXON-E framework governs the LLM through the deterministic constraints of its "Filter-then-Search" mechanism. This ensures that in the high-stakes context of legal supervision, AI outputs are reliable, explainable, and traceable, aligning with the principles of Trustworthy AI[10].

9 Conclusion and Future Work

The AXON-E framework addresses the challenge of supervising massive video archives by integrating a high-throughput pipeline with a novel knowledge-centric reasoning core. By leveraging an "Eight-Dimensional Coordinate System" and an "Advanced-RAG" mechanism, it achieves deep, precise, and automated compliance audits at scale, advancing AI application in this domain from shallow perception to deep cognition.

Future work will focus on optimizing model efficiency and expanding the knowledge base to cover more enforcement types. The ultimate goal is to transform AXON-E from a supervisory tool into a strategic instrument for data-driven governance and policy refinement.

References

[1] Cheng, H., et al. (2024). "A Survey on Large Multimodal Models for Video Understanding." arXiv preprint arXiv:2407.03159.
[2] Lewis, P., et al. (2020). "Retrieval-augmented generation for knowledge-intensive NLP tasks." Advances in Neural Information Processing Systems, 33.
[3] Gao, Y., et al. (2024). "Retrieval-Augmented Generation for Large Language Models: A Survey." Communications of the ACM, 67(6), 65-75.
[4] Yao, Z., et al. (2024). "RAG vs. Fine-tuning: A Comprehensive Survey on Knowledge Injection in LLMs." arXiv preprint arXiv:2407.06993. (Discusses hybrid search and metadata filtering as key RAG strategies).
[5] Arslanian, H., & Fischer, F. (2022). The Future of Finance: The Impact of FinTech, AI, and Crypto on Financial Services. Palgrave Macmillan.
[6] Xi, Z., et al. (2023). "The Rise and Potential of Large Language Model Based Agents: A Survey." arXiv preprint arXiv:2309.07864.
[7] Radford, A., et al. (2023). "Robust speech recognition via large-scale weak supervision." Proceedings of the 40th International Conference on Machine Learning.
[8] Bredin, H., et al. (2020). "Pyannote.audio: neural building blocks for speaker diarization." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
[9] Engin, Z., & van der Torre, L. (2024). "AI-supported governance: A new paradigm for public administration." Data & Policy, 6, e12.
[10] Jobin, A., Ienca, M., & Vayena, E. (2023). "The global landscape of AI ethics guidelines." Nature Machine Intelligence, 5(9), 974-985.