An open source collective for inter-agent collaboration

Intelligent video monitoring: How multi-agent AI systems improve flexibility and speed

Authored by Maksym Shchebuniaiev and Alona Nesterenko

Many businesses today use intelligent video monitoring. Real-time video analytics help them make better decisions, work efficiently, and ensure safety. The ability to extract actionable insights from video streams — autonomously and at scale — unlocks new possibilities for explainable, responsive, and scalable video intelligence.

As video intelligence moves to multi-agent systems, agent orchestration is a big challenge. It's less about analyzing visuals and more about coordinating task-specific agents that operate in real time. This is especially complex in an open source environment, where agent diversity and integration issues impact the system. This is where the AGNTCY, an open source collective for inter-agent collaboration, provides a transformative approach by distributing intelligence across independent, task-specific agents, enabling real-time adaptability and scalable solution — without disrupting the entire system.
 In this article, we will cover its benefits, implementation challenges, and industry-specific use cases.

Business Impact

Video monitoring with agents improves how you manage and extract value from visual data.

Benefit	Description	Impact
Modular multi-agent design	Each agent handles a specific task	Simplifies debugging, accelerates iteration, and enhances system clarity
Plug and play modularity	New agents can be added or replaced independently	Reduces integration costs and speed up innovation
Real-time interactivity	Agents respond in near real-time	Enhances operational efficiency
Reusability	Agents described via JSON files can be repurposed across domains and projects	Cuts down development time and shared infrastructure
Transparent intelligence	Each agent's logic is separated and traceable within the orchestration framework	Increases trust and compliance readiness

Breaking down complex logic into specialized agents allowed us to move faster, debug easier, and adapt to new requirements with minimal overhead.

However, while the benefits are compelling, this architectural shift introduces its own set of challenges.

Challenge	Description	Impact
Agent orchestration complexity	Coordinating agents across tasks can introduce execution dependencies	Requires robust scheduling and error-handling logic
Increased integration overhead	Inter-agent communication and message passing	Adds architectural complexity and potential bottlenecks
Data privacy risks	Decentralized systems may expose more interfaces to vulnerabilities	Requires strict access control and monitoring mechanisms
Model maintenance	Each agent may need updates over time	Adds operational overhead

As intelligent video monitoring matures, AGNTCY's modular, agent-based design is proving useful across industries. By decoupling core video processing tasks into independent agents, you can tailor systems to specific needs while maintaining performance, flexibility, and explainability.

Industry

Manufacturing

Use Case

Supporting automated product inspection, identifying defects on assembly lines, and monitoring equipment and worker safety in hazardous environments.

Industry

Retail

Use Case

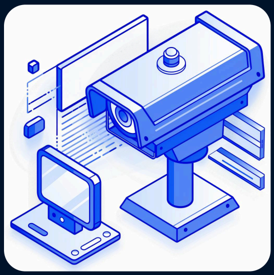
Tracking customer foot traffic, measuring engagement near product displays, detecting shoplifting, and optimizing store layouts with heatmaps and movement data.

Industry

Healthcare

Use Case

Enhancing diagnostic accuracy by detecting anomalies in medical imaging and enabling patient behavior monitoring to detect falls, distress, or prolonged inactivity.



Industry

Security and surveillance

Use Case

Detecting unauthorized intrusions, loitering, or abnormal activity through continuous analysis of live feeds. AI agents flag unusual patterns and instantly notify human operators.

Industry

Smart cities and transportation

Use Case

Managing traffic flow, detecting road congestion or illegal crossings, and ensuring crowd control during large public events using pedestrian and vehicle classification.

Modular design: Faster, flexible, and future proof

We designed a responsive, modular video monitoring system using AGNTCY's architecture as the foundational infrastructure. This let us split complex tasks into specialized agents. Each agent handled a specific job. We built an orchestration layer using AGNTCY's principles of decentralization, message passing, composability, and context awareness. This layer ensures real-time responsiveness and keeps modules clear and isolated. AGNTCY also standardizes and enables composability in multi-agent systems, providing integration protocols to connect with agents.

- Agent Connect Protocol (ACP): A standard interface and client/server SDKs to invoke agents or agentic applications.
- Secure Low-Latency Interactive Messaging (SLIM): A messaging layer that offers secure network-level communication services.

All agents in our implementation are described using Open Agent Specification Format (OASF) JSON manifests. These manifest files contain metadata about each agent's inputs, outputs, capabilities, and communication parameters. This consistent and standardized approach enables rapid onboarding, plug-and-play reuse across projects, seamless interoperability, and streamlined monitoring and managing across diverse systems.

Our implementation leverages AGNTCY's ACP and agent-to-agent communication mechanisms. ACP provides a standardized, JSON-based messaging protocol enabling seamless interactions among diverse agents, regardless of the frameworks they use. Applying ACP helps us to standardize communication between our Video Question Agent (VQA) and downstream components like the Knowledge Base Agent and SQL Agent, reducing the complexity of ad hoc integrations and preventing runtime communication mismatches. Additionally, utilizing ACP and streamlined agent-to-agent messaging through SLIM significantly enhances scalability and reduces integration overhead in our intelligent video monitoring solution.

By adopting the AGNTCY's components, we've fixed three challenges typical for multi-agent systems:

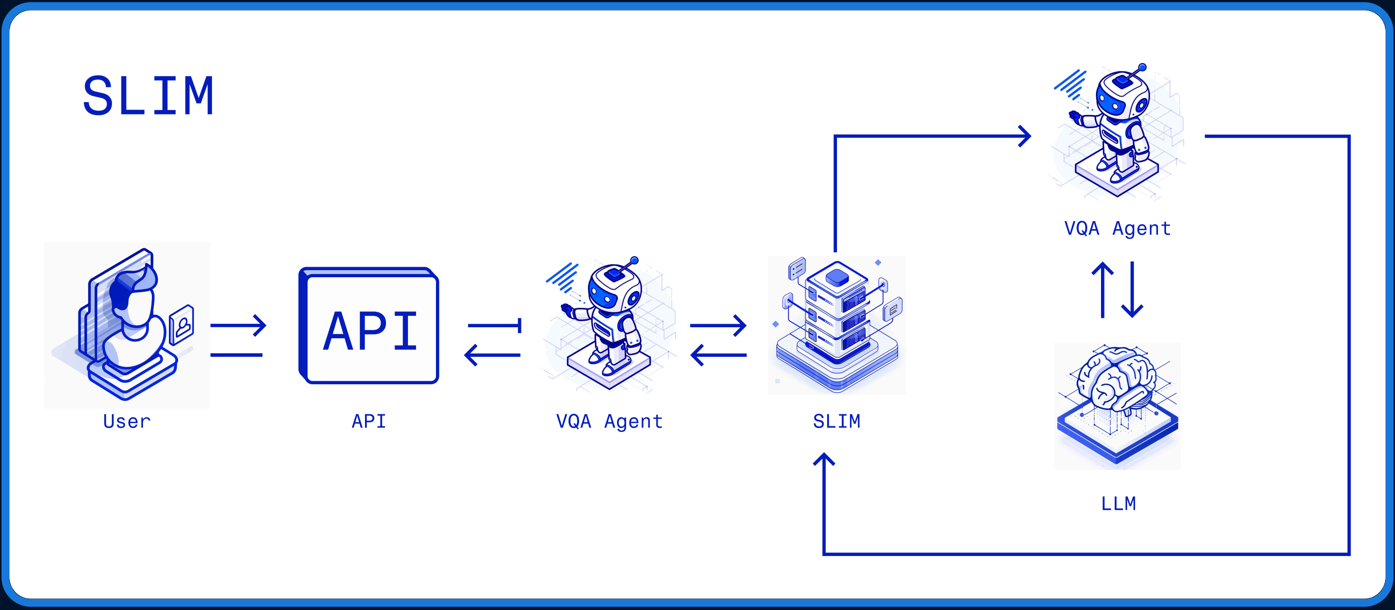
1. Complex orchestration
2. Inconsistent communication standards
3. Tight coupling between components

For example, agent-to-agent interactions such as result merging or query routing now happen through a shared protocol layer, eliminating ad hoc interfaces and minimizing the risk of runtime mismatches.

The result? A highly modular, resilient, and scalable solution that can quickly incorporate new agents or adapt to evolving workflows without disrupting the existing architecture.

To enable lightweight and scalable agent collaboration, we integrated SLIM. The diagram below shows a simplified example from our implementation, where user sends a query via the API, handled by the VQA. VQA allows users to ask questions and receive answers based on video footage. This concept was described also in our previous papers.

In our implementation, VQA serves as the entry point for user queries and orchestrates downstream reasoning tasks by leveraging AGNTCY components and services. This structure allows the VQA pipeline to dynamically adapt to different questions.



1. **SLIM** provides secure network-level communication.
2. **Question Classifier Agent** processes the intent of the query.
3. **Large Language Model (LLM)** provides enrichment or classification help.

All communication is handled via JSON-based messages encapsulated in the queue. Incorporating both approaches let us evaluate their complementary strengths and ultimately adopt a hybrid model that improves agent collaboration.

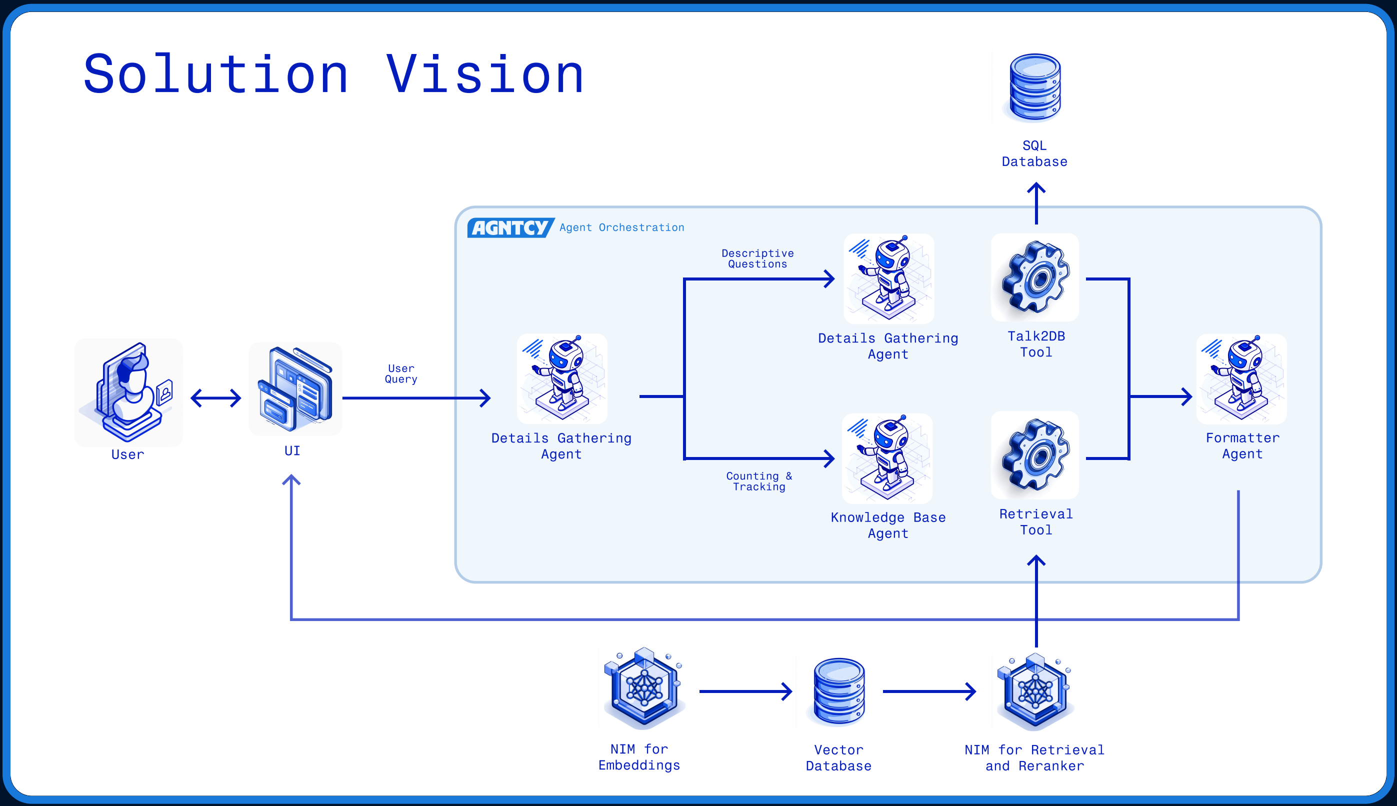
Lessons learned

Through hands-on deployment, we discovered:

1. The importance of standardized messaging (via JSON and OASF manifests) for agent interoperability.
2. The need for hybrid orchestration models to balance flexibility with control.
3. That plug-and-play architecture significantly reduces onboarding time and supports rapid iteration.

Build smarter video monitoring with modular AI agents

The architecture diagram below shows our final intelligent video monitoring application built using the AGNTCY components and services. It shows how modular AI agents, connected through standardized protocol, collaborate to deliver scalable, real-time video analytics across complex operational environments. It also provides the glue to connect diverse AI agents into cohesive workflows, allowing developers to use what's already out there — instead of reinventing the wheel.



The architecture is built around:

- Details gathering agent: Responsible for refining the user's query, requesting clarifications when needed, and routing it to the right agents.
- SQL agent: Generates structured queries using Talk2DB, interacting directly with SQL databases when descriptive data is needed.
- Knowledge base agent: Interacts with the vector store to retrieve relevant video captions and embedded scene descriptions using retrieval-augmented generation (RAG). Handles unstructured knowledge retrieval using vector databases and NVIDIA's NIM for Retrieval and Reranker models that are delivering accurate answers quickly at scale.

- Formatter agent: Collects results from SQL and retrieval agents, then unifies and formats the final response for output.
- Thanks to AGNTCY's plug-and-play architecture and standardized messaging, we can reuse agents across workflows and extend functionality without disrupting the system. Its modular structure allows functionality to be extended—for example, to support alternative output formats or regulatory requirements—simply by chaining it with other agents. This flexibility ensures that enhancements or changes can be introduced incrementally. That said, our solution did reveal a few practical limitations: While AGNTCY presents a well-defined conceptual model, its documentation and SDK examples are still evolving, which may require more effort during initial adoption.
- Minor inconsistencies and occasional dependency issues can introduce slow-down during development and onboarding, particularly for new contributors.
- Continuous monitoring of the ecosystem is needed, as rapid evolution in LLM tooling and frameworks may affect compatibility.

Despite these challenges and thanks to AGNTCY's flexible interfaces, developers can test individual agents in isolation, simulate inter-agent communication, and then deploy them in a unified pipeline.

Improve multi-agent systems with AGNTCY

At SoftServe, we transform advanced technologies into practical, scalable solutions, and our adoption of AGNTCY's components and services for our intelligent video monitoring exemplifies this approach. AGNTCY enabled us to move beyond monolithic, tightly coupled pipelines to a flexible, modular system that maintains real-time performance without sacrificing transparency or control. Looking ahead, AGNTCY opens new opportunities beyond video analytics—for example, empowering business users with conversational data access through solutions like "Talk2DB," enabling personalized retail experiences via multimodal AI assistants, and streamlining internal workflows with agents for documentation, analytics, and data visualization. By driving AGNTCY components into these strategic areas, SoftServe continues to lead in delivering intelligent, future-ready agentic solutions across industries.

Visit agntcy.org to explore how Outshift by Cisco is building the Internet of Agents, the collaboration layer that will let AI agents discover each other and work together.