



ASK THE EXPERT

The Cloud Isn't the Default Anymore: What's Driving the Shift to Edge AI in Video Security



DAVID RASKE

Director of safety and security and smart cities within Intel Corporation's Edge Computing Group

Video security systems are increasingly built across edge, cloud, and hybrid architectures that shape how and where video data is processed and analyzed. These deployment choices are driven by a variety of factors including latency, cost, and data sovereignty as well as the expanding role of AI at the edge.

David Raske, director of safety and security and smart cities within Intel Corporation's Edge Computing Group, explains the state of edge, cloud, and hybrid video security architectures and what's next for the industry.

What are you seeing in the market when it comes to edge, cloud, and hybrid architectures for video security deployments?

We're seeing an expansion of different architectures, which reflects the diverse business requirements organizations have and the increased flexibility of today's technology solutions to meet those needs. Whether the priority is cost, performance, or other operational factors, integrators and end users have more choices than ever.

What's also interesting is that we're seeing movement from both sides of the market. Traditional edge solution players are now offering cloud options, while some of the previous cloud-only disruptors that initially promoted deployments without edge devices are now offering them as part of their solutions. There's been a pendulum swing on both sides.

What types of environments or use cases are best suited for edge AI video analytics?

The case for video analytics at the edge comes down to five key factors. The first is latency: how quickly you can respond to an observed event. The second is the total cost of ownership, including both upfront hardware costs and long-term operational costs. Third is bandwidth constraints, whether total bandwidth needed to stream video or reliability of that bandwidth. Fourth is data sovereignty and privacy, since some organizations aren't comfortable sharing data outside their own walls. And fifth is regulatory compliance, because some industries have restrictions around what can and cannot be done in the cloud.

So performance, reliability, economics, and regulation are all driving edge adoption.

Sponsored by



Edge is also a flexible concept. Depending on who you ask, the definition can vary. For us, it includes everything from ruggedized systems in specialized environments to traditional on-prem systems deployed in offices or server-rack environments outside of massive data centers.

From an AI inference perspective -- running the models rather than training them -- the edge is highly capable for video analytics and video security workloads. Once deployments are tested against real-world conditions, performance tends to be predictable, reducing the need for the cloud's dynamic scalability. We don't see many limitations with edge analytics today.

Where does cloud architecture continue to offer advantages over hybrid and edge deployment models?

If you split AI into two areas, training and inference, almost all model training today still happens in the cloud because of the massive infrastructure requirements. In our industry, though, we tend to care more about the inference side.

Where the cloud can still provide significant value is in aggregating data across multiple physical locations to create a centralized view. That can be done in other ways, but it's typically easiest in cloud-based architecture.

The cloud is also well-suited for video analytics workloads that rely on analyzing data over time rather than responding to events in real time. Use cases like heat mapping or trend analysis are good examples.

What are the biggest misconceptions organizations still have about edge, cloud, and hybrid AI video security architecture?

One of the biggest misconceptions is that you need cloud to run AI, especially generative AI. While training typically happens in the cloud, the models themselves can run at the edge.

Another misconception is that AI deployments are super expensive and require dedicated GPU cards. GPUs are important for training, but inference workloads can often run on CPUs "only", which includes CPUs with integrated accelerators. As AI continues to become more affordable to deploy, we expect adoption to accelerate even faster.

A third misconception is that edge AI or hybrid AI architectures are difficult to implement. We've focused heavily on simplifying that process through technologies like Intel's Open Edge Platform and Metro AI Suite, which include open source software and inference tools that help organizations develop and deploy AI more quickly. Even transitioning from a proprietary AI software stack to an open source stack like Intel supports is no longer as difficult.

How do you expect edge AI and hybrid video security architectures to evolve over the next three to five years?

We see a couple of key trends. One is that edge AI technology -- like compute more broadly -- is improving rapidly, with significant gains in capability at a lower cost. As a result, we'll continue to see increasing AI capability embedded across almost all types of edge systems.

While some devices today are still power-constrained with limited performance, we expect that, across the board, edge form factors will become significantly more capable over time. Some of this will come from continued innovations in power-efficient acceleration, like the NPU found in Intel® Core™ Ultra processors.

At the same time, models are becoming more optimized and, in some cases, smaller, creating a powerful combination: greater capability running on lighter workloads across both edge and hybrid architectures.

Another important trend is the rise of physical AI at the edge, whether it's robotics, drones, or other AI-controlled systems. In video security, that extends to things like access control, lockdowns, and other automated physical or guard response via agentic AI.

Over time, we'll also see more generative AI-based video analysis and new solution designs that take advantage of all available compute -- edge, cloud, or hybrid. The industry is increasingly focused on dynamically using available compute to deliver value.

Ultimately, edge AI will become more mainstream in all industries, driven by practical performance benefits and clear economic advantages.