Google Cloud sertn

Verifiable computer vision.

Sertn on Google Cloud Confidential Space.

July 2026

Verified computer vision at production scale.

Sertn is an end-to-end computer vision platform that takes organizations from raw data to deployed, production-grade models. Every prediction produces a publicly verifiable proof, which effectively resolves the 'Trust Deadlock': the fundamental conflict between protecting proprietary IP and providing independent auditability that has historically stalled AI adoption in regulated industries.

To make this possible across operationally sensitive and regulated environments, Sertn built its Proof of Inference architecture on Google Cloud Confidential Computing, deploying inside Confidential Space to guarantee model confidentiality, runtime integrity, and independent verifiability across organizational boundaries. Delivering this caliber of cryptographic trust is only achievable because it is built upon the foundational security and industry-leading reputation of Google Cloud.

The Sertn platform builds on more than seven years of real-time computer vision deployment experience in airport infrastructure, with over 1 billion production proof artifacts generated to date. This includes more than 900,000 verifiable proofs at a single international airport deployment, with active expansion to a second major North American hub.

The challenge.

Enterprise computer vision is increasingly responsible for operational decisions in environments where errors carry financial, regulatory, or safety consequences. Airports coordinate vehicle movement through it. Industrial facilities monitor restricted zones. Logistics systems track physical assets across multi-vendor environments. Hospitals are deploying automated monitoring into patient-facing infrastructure.

Across these environments, the constraint on deeper deployment has shifted from model capability to a question of trust between the parties involved. Model owners are reluctant to deploy proprietary weights into infrastructure they do not control. Operators are reluctant to accept models they cannot audit. Regulators no longer treat vendor-controlled telemetry as sufficient evidence. Multi-vendor environments lack any participant willing to serve as the neutral authority for the others.

This condition, **the Trust Deadlock,** has four sources:

Model IP risk.

Model owners are unwilling to deploy proprietary weights to environments operated by parties they do not control, since cloud operators, infrastructure administrators, and customer staff all sit inside the conventional trust boundary in ways a legal contract cannot constrain.

No neutral authority.

Airports, logistics hubs, and industrial sites combine infrastructure owners, vendors, contractors, and regulators inside the same physical environment, and no single participant can credibly arbitrate operational events involving all of them.

Customer data risk.

Operators are unwilling to feed sensitive data into vendor-controlled systems producing decisions they cannot independently audit, particularly when the cost of an incident sits with the operator rather than the vendor.

Vendor-controlled audit trails.

Regulators and auditors require more than telemetry generated by the party under review, recognizing that vendor-curated logs are a record the vendor agrees to disclose, not independent evidence.

Conventional architectures resolve this by concentrating trust in a single actor. That arrangement does not survive adversarial review in regulated computer vision. Sertn needed an architecture where confidentiality and independent verification coexist, without forcing any participant to expose what they protect.

The solution.

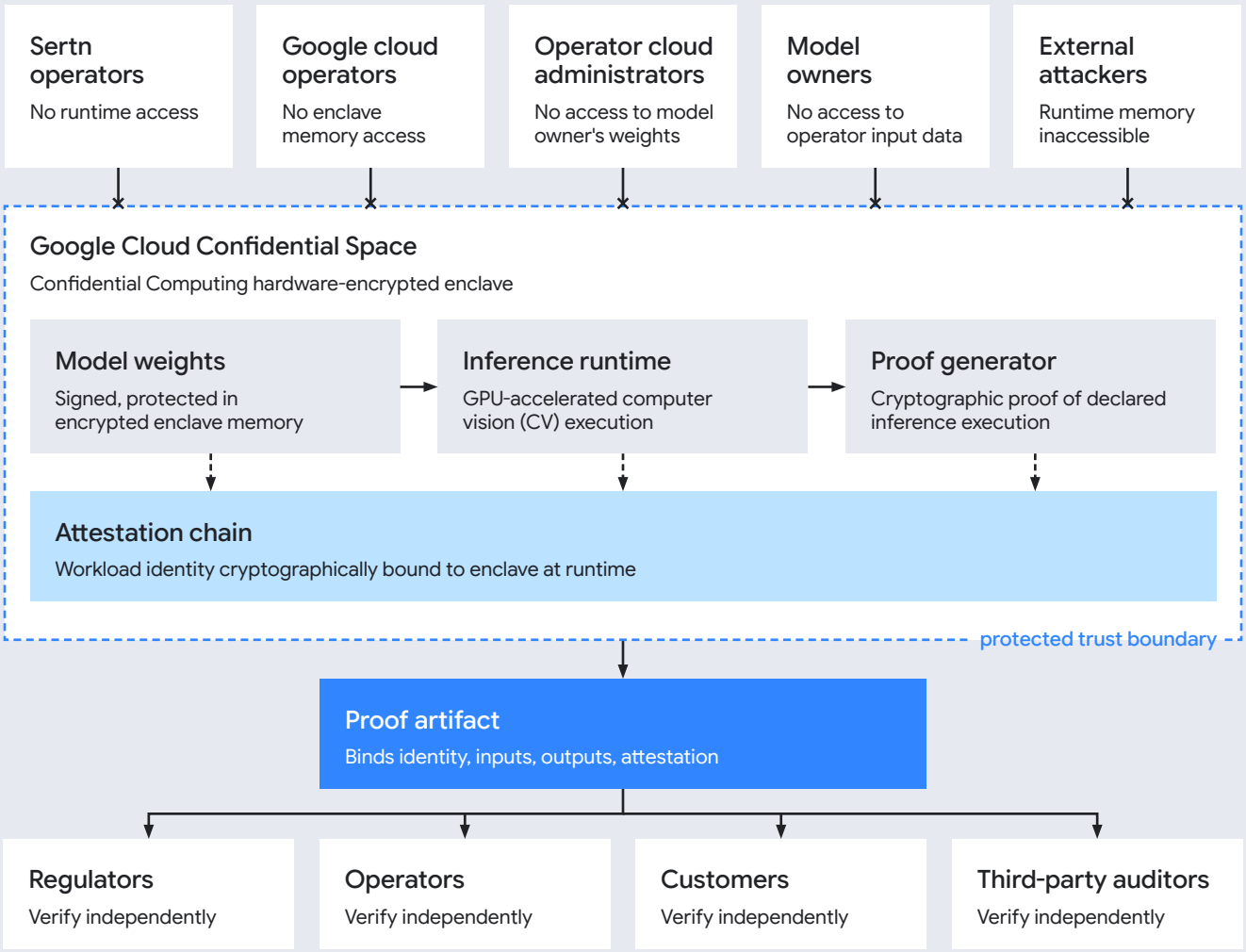
Sertn × Google Cloud Confidential Compute

Sertn deployed its Proof of Inference architecture inside Google Cloud Confidential Space, a trusted execution environment (TEE) powered by AMD Secure Encrypted Virtualization (SEV) or Intel TDX.

The trust boundary

Deployment architecture

The architecture protects two parties from each other. Operator admins cannot read the model owner's weights. Model owners cannot read the operator's input data. Adversaries do not cross into the enclave. Proof artifacts cross out.

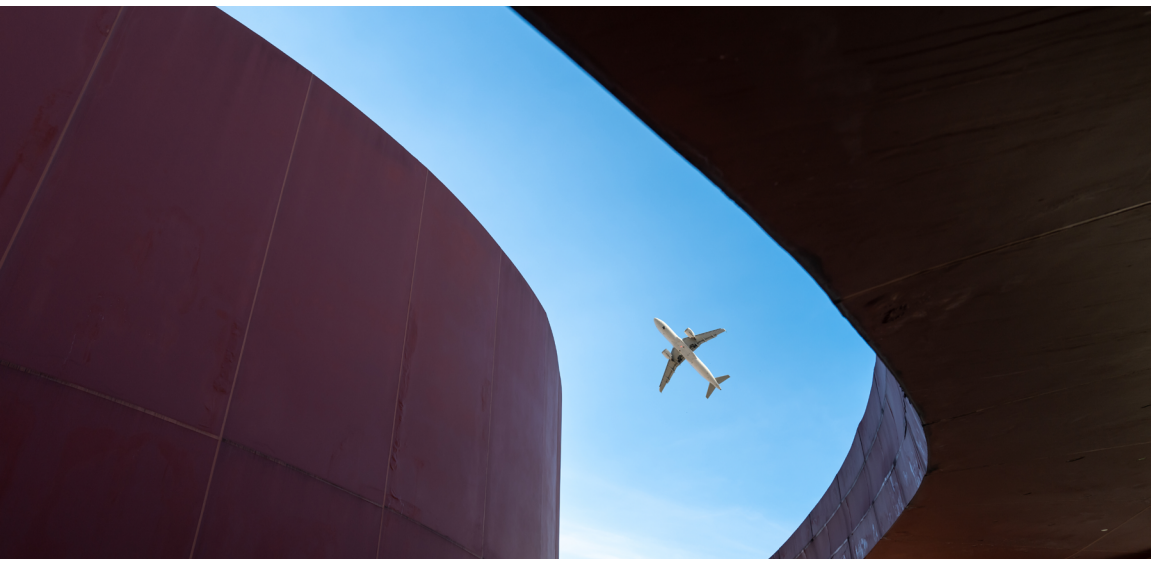


Inside the confidential workload boundary, model weights remain encrypted in memory and inaccessible to the cloud operator, the customer's infrastructure administrators, and Sertn's own personnel. Inputs are protected from the model owner outside the enclave. Workload identity is established through Google Cloud's attestation chain, cryptographically tied to the confidential environment.

Proof of Inference[1] extends this baseline by attaching cryptographic evidence directly to the inference process. A proof artifact binds the deployed model identity, the declared inputs, the runtime context, the generated outputs, and the workload attestation into a single verifiable record. The proof is generated during execution, not reconstructed afterward from logs.

This deployment pattern is in production. Sertn operates at TRL 8 to 9 (*Technology Readiness Level, a maturity scale used in regulated and defence procurement, where 8 to 9 indicates fielded operational deployment*[2]), with more than 1 billion proof artifacts generated across production workloads, including a live deployment at an international airport.

Confidential Computing secures these workloads with under 5% latency overhead, and attestation costs are amortized across requests via runtime token caching.



1. arXiv:2508.06972 | 2. arXiv:2510.21024

How it works.

The protected boundary excludes six adversary classes that conventional computer vision deployments cannot simultaneously exclude.

Sertn personnel.

Vendor operations, engineering, and administrative staff have no runtime access. Shell access, debugger attachment, and memory inspection are blocked at the enclave boundary.

Google Cloud personnel.

Cloud operators, SREs, and infrastructure administrators cannot read model weights or inference data in plaintext, because Confidential Computing enforces memory isolation in hardware rather than through policy.

Customer cloud administrators.

Customers operating the deployment infrastructure cannot extract model IP or inspect inference internals, since the confidential workload is isolated from the surrounding environment.

External attackers with infrastructure access.

Adversaries who compromise the surrounding host encounter encrypted memory, so the workload boundary holds despite host-level compromise.

Supply-chain attackers.

Tampered machine images cannot execute, because signed workload identity is verified at instantiation and attestation reports bind execution to the declared software version.

Malicious co-tenants.

Per-workload memory encryption keys enforce isolation in hardware, so co-tenant compromise does not yield access to a confidential workload's memory, weights, or proof generation.

The verifier is no longer constrained to the deployment vendor. Regulators, operators, customers, partners, and third-party auditors verify the same proof artifact against the same public keys.

Verification flow.

Verification is built into the execution path itself. First, Sertn's execution environment is an extension of a hardened base image to limit the attack surface. As a workload is instantiated inside Confidential Space, Google Cloud's attestation chain establishes both the integrity of the image and the VM itself before any data enters the boundary. Signed model weights and inputs cross into the enclave only after secure boot and setup completes, and inference and proof generation run as a single sealed step. The proof artifact leaves the enclave; the weights and runtime stay inside.

Seven stages, one sealed execution

Verification flow

Setup

Instantiate confidential workload 01

Confidential Space enclave starts

Establish workload identity 02

Google Cloud attestation chain confirms enclave integrity

Provisioning

Provision model weights 03

Signed weights loaded into enclave, encrypted in memory

Transmit input data 04

Attestation-bound channel delivers CV input into the enclave

Execution

Execute inference and bind proof 05

Model runs while Proof of Inference binds model, inputs, outputs, attestation.

Export proof artifact 06

Artifact leaves enclave; model weights and runtime stay protected

Verification

Verify independently 07

Third parties verify against public keys, no model or data access required

Structural property

Proof is bound at the moment of execution, not assembled from logs afterward. This is what allows the artifact to survive review by parties who do not trust the operator's telemetry.

Why Google Cloud.

01 Workload identity binding.

Confidential Space cryptographically links workload identity to attestation throughout runtime, allowing Proof of Inference to anchor every artifact to a verified execution environment.

02 Global availability.

Regional deployment through Google Cloud's many global data centers supports customers who have unique data sovereignty obligations, including critical infrastructure operators and regulated computer vision deployments.

03 GPU-enabled confidential computing.

Production computer vision throughput requires GPU acceleration, and the Confidential Space GPU pathway supports inference at scale without separating protected execution from accelerated compute.

04 Mature attestation chain.

Google Cloud's attestation infrastructure integrates with cryptographic proving pipelines without bespoke integration work, making the deployment pattern viable at production scale.

What Confidential Compute enables.

By deploying on Google Cloud Confidential Compute, Sertn demonstrates:

Multi-vendor neutral environments.

At a major North American airport, multiple operational stakeholders rely on overlapping computer vision systems within shared infrastructure zones. No single vendor serves as the authority for validating events across all participants, and proof artifacts produce a shared source of truth none of the parties independently control.

Repeatable across sites.

Following production deployment at a major North American airport, the architecture is now expanding to a second major North American hub, where Sertn is building the evidence layer for existing operational systems. The deployment pattern transfers across independent infrastructure operators.

Cross-organizational AI workflows.

Model owners, infrastructure operators, and oversight bodies in different jurisdictions verify the same proof artifact without exposing their systems to each other.

Protected model IP at deployment.

Model owners retain confidentiality of proprietary weights throughout deployment while supporting independent third-party verification of execution.

Audit trails that survive adversarial review.

Evidence is cryptographic and tied to execution rather than reconstructed from vendor-controlled telemetry, so the audit trail holds because no participant under review controls the evidence source.

1B+

verifiable proof artifacts generated across production workloads

900K+

proofs at a single international airport

2nd

major North American hub expansion underway

< 5%

latency overhead using Confidential Computing

TRL 8-9

operational deployment

Looking ahead.

The collaboration between Sertn and Google Cloud continues to expand. Hardware-accelerated proving on confidential infrastructure continues to close the gap between production inference throughput and independently verifiable execution. Regulated industries beyond computer vision, including healthcare imaging, financial decision-making, and critical infrastructure monitoring, are evaluating architectures in which verification becomes part of the deployment stack rather than a downstream compliance process.

Sertn presented its Auditable AI governance approach to the SAE International G-12 Standards Committee, reflecting growing interest from standards and regulatory bodies in independently verifiable AI for safety-critical systems.

Sertn's CVPR 2026 submission on production-scale proof generation extends this work to broader deployment contexts, with adoption growing across AI infrastructure, defense, and regulated-industry partners. With Google Cloud, Sertn is moving towards the next generation of confidential computing primitives for verifiable AI.

Learn more.

Sertn builds infrastructure for Verified Confidential Inference using Google Cloud Confidential Space and Proof of Inference architecture, designed for enterprise computer vision deployments in regulated and operationally sensitive environments.

Learn more about sertn.ai and Google Cloud Confidential Space.

Google Cloud sertn