From Trust Assumptions to Trust Evidence

Confidential computing for brownfield medical devices and agentic AI.

Brian Trzupek

Senior Vice President, Product · DigiCert

digicert

CONFIDENTIAL COMPUTING SUMMIT 2026

Hosted by THE LINUX FOUNDATION + OPAQUE

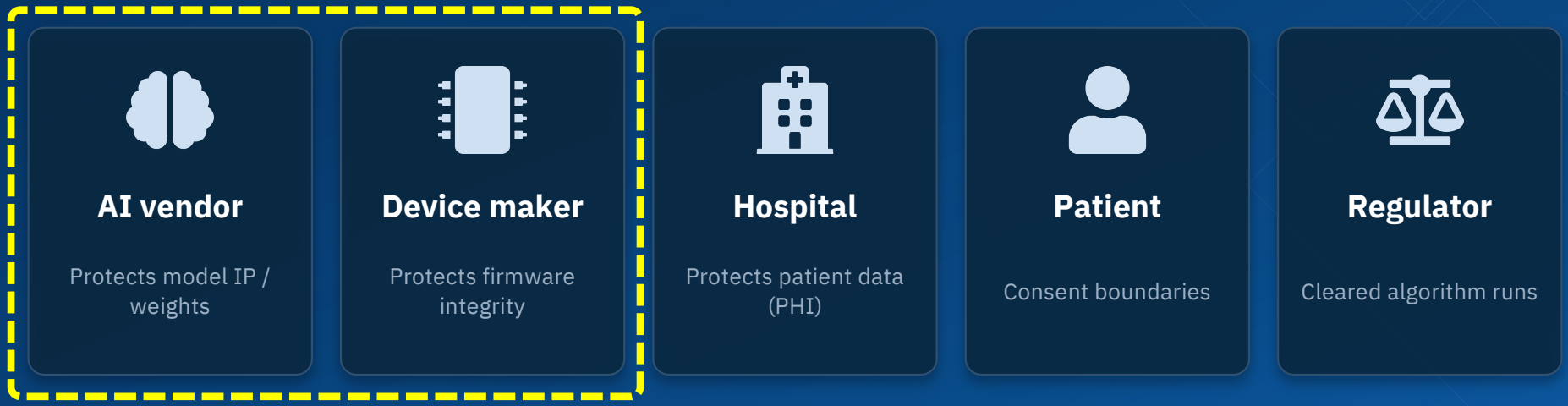
THE STATUS QUO

Parties with conflicting interests agree to trust each other procedurally — because no technical mechanism exists to verify the claims they make.

Contracts. Audits. Paper attestations. Faith.

AN MRI RUNNING AN AI DIAGNOSTIC MODEL

~~Five~~ FOUR stakeholders. Zero verification.



Every relationship between them is "take it on faith."

FROM FAITH TO EVIDENCE

Each conflicting interest maps to one CC primitive.

Stakeholder	Protects	Takes on faith today	Verified with attestation
AI vendor	Model weights / IP	Hospital won't extract the model	Encrypted weights decrypt only inside an attested TEE (memory is HW-encrypted). Attestation gates the key, not the model — host & cloud see only ciphertext.
Device maker	Firmware integrity	Device not tampered in the field	Measured boot + PCR / RIM verification proves the firmware state.
Hospital	Patient data (PHI)	Vendor won't exfiltrate scans	Data processed in an enclave invisible to the vendor, cloud operator, and co-tenants.
Patient	Consent boundaries	Scan isn't repurposed for training	TEE-enforced policy; weight export gated; tamper-evident audit.
Regulator	Only the cleared model runs	Hospital runs the approved version	Model hash in the attestation equals the cleared-version hash.

WHY CONFIDENTIAL COMPUTING…

It lets mutually distrustful parties each verify the one claim they care about — without disclosing what they protect.

Encrypted in use

Not just at rest or in transit. Data and code stay protected while being processed.

Hardware-rooted

Root of trust in silicon — not in a vendor's word or a config file.

Remotely verifiable

The proof can be appraised by a party who was never on the box.

THE SAME PROBLEM, IN RATS TERMS · RFC 9334

This five-party mess is an attestation topology.

Device	Attester
Device maker + AI vendor	Endorsers / Reference Value Providers
DigiCert	Verifier
Regulator	Relying Party
Key Broker	Relying Party (enforcing)

We didn't design this mapping.

The problem defined it. A contractual / procedural mess becomes a clean set of attestation roles.

PROBLEM 1

Brownfield medical devices

Prove the box. — Attestation at rest, down to the hardware.

BROWNFIELD IS ALWAYS THE HARD CASE

The vendors show greenfield. The customers are brownfield.

Greenfield · cloud — the easy case

Intel TDX | RTMR | DCAP quote

Provision a fresh confidential VM, capture RTMRs, pull a DCAP quote. Mostly solved — and what most CC demos show.

Brownfield · field — the real case

TPM 2.0 | Measured boot | PCR golden values

Deployed years ago. FDA-cleared as a fixed configuration. Can't be taken offline. Has a TPM — sometimes not even that. No discrete TDX.

THE RETROFIT — THREE THINGS

Add evidence without touching the cleared payload.

1 · Measured boot baseline

TPM PCRs across UEFI → bootloader → kernel → app. Capture golden values once, in a known-good state.

2 · Model integrity

SHA-256 + Cosign. Weights encrypted at rest. Key released only on a passing attestation — fail-closed.

3 · Remote attestation

TPM quote → Verifier → signed token. RFC 9334 vocabulary, end to end.

NON-TPM, Unmodified Device Path:

Leverage exiting device birth certificate (crypto), and derive golden value.

Decryption based of mutual crypto from birth certificate and customer crypto.

Primitive attestations from derived claim, proof only of ‘file state’...

HOW THE SERVICE VERIFIES

Evidence in, signed assurance out.

1

Evidence collection

Workload collects a TPM quote / TD quote using go-tpm-tools or go-tdx-guest.

2

Submit to DigiCert

Sent to the verification service via REST. Async — built for fleet volume.

3

Verify against RIMs

Check the cert chain AND the signature over measurements, cross-referenced to RIMs / golden PCR values.

4

Signed attestation

A signed token / certificate returned for use or publication.

ENFORCEMENT WITHOUT BREAKING THE CLINIC

“Tamper → model goes dark” is a feature in a Enterprise. *In an MRI machine* it can be a safety hazard.

Gate at safe points

Enforce at boot and at key-release — not mid-procedure.

Re-attest on a cycle

Continuous re-attestation detects drift between procedures.

Clinically-safe degraded mode

Define a safe fallback — not an abrupt kill while a patient is in the bore.

No discrete TPM? The retrofit bridge: **`swtpm / vTPM`** → representative of vSphere vTPM · Proxmox KVM · Azure Trusted Launch. Last fallback, derive crypto from device birth certificate.

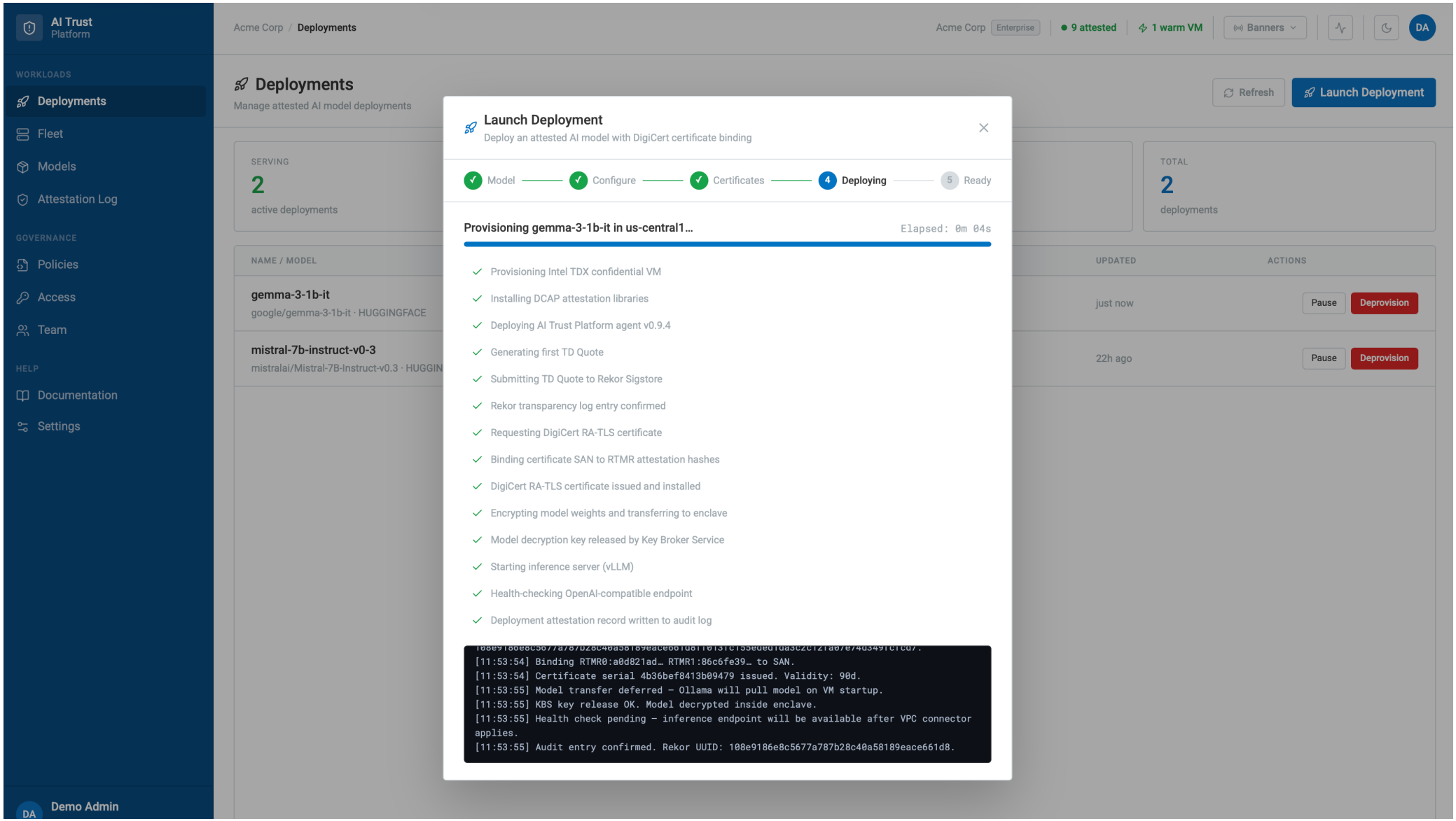
NOT SLIDEWARE

A real deployment writing a real receipt to a public log.

- Provisioning confidential VM
- Installing attestation libraries
- Deploying AI Trust Platform agent
- Generating first quote
- Submitting quote to Rekor Sigstore
- Rekor transparency log entry confirmed
- Binding certificate SAN to attestation hashes
- DigiCert RA-TLS certificate issued
- Model key released by Key Broker Service
- Deployment attestation written to audit log

```
[11:53:54] Binding RTMR0:a0d821ad… to SAN.
[11:53:54] Cert serial 4b36bef8413b09479 issued. 90d.
[11:53:55] KBS key release OK. Model decrypted in
enclave.
[11:53:55] Audit entry confirmed.

Rekor UUID:
108e9186e8c5677a787b28c40a58189eace661d8
```


AI Trust
Platform
WORKLOADS
Deployments
Fleet
Models
Attestation Log
GOVERNANCE
Policies
Access
Team
HELP
Documentation
Settings
Demo Admin
Acme Corp / Deployments
Acme Corp Enterprise
9 attested
1 warm VM
Banners
DA
Deployments
Manage attested AI model deployments
Refresh
Launch Deployment
SERVING
2
active deployments
TOTAL
2
deployments
NAME / MODEL
UPDATED
ACTIONS
gemma-3-1b-it
google/gemma-3-1b-it · HUGGINGFACE
just now
Pause
Deprovision
mistral-7b-instruct-v0-3
mistralai/Mistral-7B-Instruct-v0.3 · HUGGIN
22h ago
Pause
Deprovision
Launch Deployment
Deploy an attested AI model with DigiCert certificate binding
Model
Configure
Certificates
4
Deploying
5
Ready
Provisioning gemma-3-1b-it in us-central1…
Elapsed: 0m 04s
Provisioning Intel TDX confidential VM
Installing DCAP attestation libraries
Deploying AI Trust Platform agent v0.9.4
Generating first TD Quote
Submitting TD Quote to Rekor Sigstore
Rekor transparency log entry confirmed
Requesting DigiCert RA-TLS certificate
Binding certificate SAN to RTMR attestation hashes
DigiCert RA-TLS certificate issued and installed
Encrypting model weights and transferring to enclave
Model decryption key released by Key Broker Service
Starting inference server (vLLM)
Health-checking OpenAI-compatible endpoint
Deployment attestation record written to audit log
[11:53:54] Binding RTMR0:a0d821ad… RTMR1:86c6fe39… to SAN.
[11:53:54] Certificate serial 4b36bef8413b09479 issued. Validity: 90d.
[11:53:55] Model transfer deferred — Ollama will pull model on VM startup.
[11:53:55] KBS key release OK. Model decrypted inside enclave.
[11:53:55] Health check pending — inference endpoint will be available after VPC connector applies.
[11:53:55] Audit entry confirmed. Rekor UUID: 108e9186e8c5677a787b28c40a58189eace661d8.

PROBLEM 2

AI agent identity

Prove the actor. — Attestation in motion, across a software mesh.

AGENTS ARE WORKLOADS, NOT USERS

Traditional IAM was built for humans. Agents break them.

The gap

No standard way to identify an agent, verify its execution environment, or know who authorized it. OAuth assumes a human; the fallback is static API keys — we've all spent time removing these.

The standards have converged

Gartner | IETF WIMSE | NIST NCCoE

The community's conclusion: this is a workload identity problem — known, trusted, governed — not a human IAM bolt-on.

Read the Digicert NIST NCCoE white paper.

THE AI AGENT PASSPORT · OPEN SPEC, CONTRIBUTED TO NIST NCCOE

A signed credential, co-issued with the SVID. JWS / ES256.

1 · Identity

SPIFFE binding · owner · trust domain

2 · Policy

Capability grants · data-class ceiling · allowed peers · kill flag

3 · Lineage

Model provenance · delegation chain · supply-chain evidence

4 · Ownership

Data-handling constraints · governance refs · A2A AgentCard-compatible

Expandable by design, for other attributes

THE HINGE BETWEEN THE TWO PROBLEMS

The Passport's Lineage layer carries the attestation forward.

PROVE THE BOX

TEE attestation

Rekor UUID 108e9186…661d8

→

PASSPORT · LAYER 3

Lineage embeds the receipt

→

PROVE THE ACTOR

Travels with the agent

Vertical trust — rooted in silicon — projected into a horizontal, portable credential.

THE THROUGH-LINE

One continuous chain. Every link is evidence, not faith.

WHERE THIS COMMUNITY COMES IN

The chain has gaps only standards work can close.

Reference values for AI artifacts

RIM / reference-value formats for weights and adapters — like we have for firmware.

RATS profiles that span layers

Profiles that reach across the device and the workload boundary, not stopping at the device.

WIMSE-for-agents

Workload identity that lands cleanly on probabilistic agents.

SCITT for the audit chain

Transparent, append-only records across the whole lifecycle.

DigiCert is **contributing**, not dictating — positions already filed with NIST NCCoE.

We're building this today.

From trust assumptions to trust evidence. If any of these problems sound like yours — come talk to us.

Download our AI Trust Architecture White Paper ->

Deploying & Protecting a Sensitive AI Model with DigiCert

End-to-end: model selection → tamper-proof deployment → controlled access → authorized fine-tuning → automated kill switch

Healthcare / FDA SaMD | Financial Services | Defense / Intelligence | Legal / Pharma R&D

1 Source Model

MODEL ORIGIN

Select from **HuggingFace**, internal registry, or partner-provided weights.

The asset: A trained model with proprietary or regulated value — weights, architecture, fine-tuning data.

Examples:
Clinical diagnostic model
Quant trading strategy
Legal contract analyzer

2 Register & Sign

IDENTITY & INTEGRITY

SHA-256 hash — cryptographic fingerprint of the model weights.

Digital signature — DigiCert-anchored chain of custody.

Model is **encrypted** before storage. Only attested environments can decrypt.

Trusted Model Registry

sha256: a3f19c22bde1…
cosign: ✓ verified

3 Confidential Compute

TDX · TPM · vTPM · NVIDIA · MORE

DIGICERT MANAGED TEE

Model runs inside a **Trusted Execution Environment** — encrypted in memory, isolated from host OS, invisible to cloud provider. Supports **Intel TDX, TPM, vTPM, NVIDIA,** and more.

Continuous attestation verifies environment integrity.

Decryption keys **only released** after attestation passes.

In Cloud | On Device

4 Controlled Access

Inference Users

Clinicians, apps, API consumers query the model. **Cannot access weights.** Policy-enforced read-only.

Fine-Tuner Users

ML team adjusts weights **inside the TEE**. Every change produces a **new signed version**. Old version preserved. Full audit trail.

Re-signs & re-registers new version

Other Users

Streaming, **embedding** & weight exporting are policy **protected.**

5 Monitor & Respond

CONTINUOUS ATTESTATION

Platform re-attests the TEE every cycle. If measurements drift — **attestation fails automatically**. Key release stops. Model goes dark.

Kill Switch

Automated: Tamper detected → keys revoked → model offline instantly.

Manual: Operator can suspend any workload from the dashboard at any time.

Tamper-Evident Audit Trail

Every event across all stages — registration, attestation, key release, fine-tune, retirement, kill switch — is recorded as an immutable, cryptographically-signed receipt on a transparency log (Sigstore Rekor). Auditors can independently verify any claim.

HIPAA | FDA SaMD | SOC 2 | NIST AI RMF