digicert®

Trust Without Compromise in AI

How Enterprises Can Evolve To Prepare for the AI Revolution

WHITE PAPER

Executive Summary

Artificial intelligence is no longer experimental—it's now a foundational component of enterprise strategy.

From generative models that produce content to autonomous agents making operational decisions, AI is being woven into the fabric of digital business.

But as capabilities evolve, so do threats. AI models are distributed via opaque supply chains. Synthetic media can be produced at scale. Autonomous agents increasingly act with limited oversight. And regulators are raising the bar on transparency, documentation, and accountability.

This paper defines DigiCert's perspective on how enterprises can evolve their trust infrastructure to meet this moment. You'll learn the essential capabilities required to prove the authenticity and integrity of AI systems, content, and outcomes—built on the same cryptographic foundations that have powered digital trust for decades.

Why the Status Quo Fails in the Age of AI

Trust frameworks were designed for predictable, deterministic systems. Traditional software behaves consistently, with clear versioning and code provenance.

AI systems, by contrast, are dynamic and opaque:

- Models evolve without clear audit trails.
- Training data may be proprietary, incomplete, or unverifiable.
- Outputs are probabilistic and harder to validate deterministically.

Consider these real-world scenarios:

A pre-trained model sourced from a public repository is integrated into a critical application. Months later, a vulnerability is discovered in its training data, but there's no record linking the dataset to your deployment.

A synthetic document, generated by an AI agent, is submitted as evidence in a contractual dispute. It looks legitimate but lacks cryptographic proof of authorship.

An autonomous agent executes transactions without an associated credential lifecycle, making it impossible to revoke access if the model is compromised.

In each case, trust breaks down because traditional controls weren't designed for AI's unique properties.

What this means: Organizations must rethink how they verify, attest, and monitor models and content. Trust must become provable and continuous—not assumed.

Principles for AI-Ready Trust Infrastructure

Based on our research and customer dialogues, DigiCert believes four principles are foundational:

Provenance Before Adoption:

You must be able to trace the lineage of any model or content before you can trust or deploy it.

Lifecycle Governance Over Static Certification:

Trust cannot be a one-time event. It must adapt as models evolve, credentials rotate, and regulations change.

Cryptographic Assurance Over Metadata:

Model cards, bills of materials (BOMs), and credentials are only valuable if tamper-evident and cryptographically signed.

Human Oversight Over Autonomous Judgment:

Even the most advanced AI must operate with clear lines of human accountability.

These principles underpin the emerging trust capabilities that every enterprise will need.

From Theory to Practice: The Capabilities that Matter

Model Signing: Establishing the Baseline of Integrity

The Problem:

Without a verifiable signature, no model can be reliably trusted. Unsigned or improperly signed models create an open door for tampering, license violations, or the introduction of malicious weights.

The Capability:

Model signing applies proven cryptographic techniques—hashing, PKI-issued certificates, and tamper-evident signatures—to AI artifacts. Each model's signature can include:

- Unique identifiers and version numbers
- Hashes of model weights and configuration files
- Issuing organization credentials

Example in Practice:

A machine learning operations (MLOps) pipeline uses DigiCert's signing service to automatically sign TensorFlow models during build. The signature binds the model to a known version of the training dataset and configuration. Downstream systems can verify the signature before deployment.

Model Card Provenance and AI BOMs: Making the Opaque Transparent

The Problem:

Even a signed model can hide significant risk if you don't know how it was trained or what limitations it has.

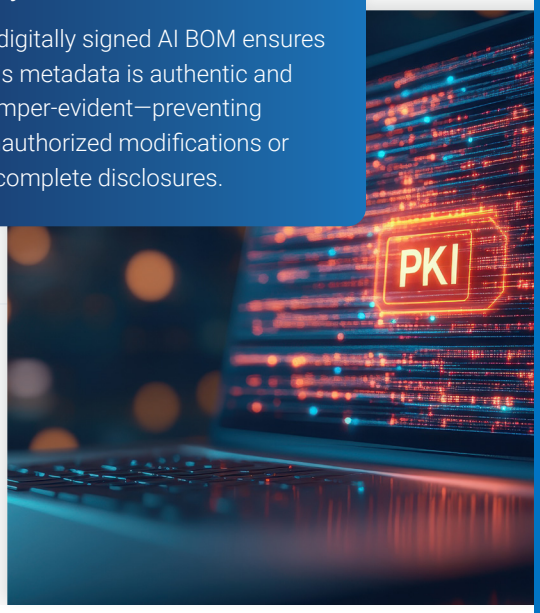
The Capability:

Model card provenance (MCP) and AI bills of materials (BOMs) create a structured, machine-readable record of:

- Training data sources
- Licensing and usage restrictions
- Known biases or constraints
- Compliance attestations

Why PKI Matters:

A digitally signed AI BOM ensures this metadata is authentic and tamper-evident—preventing unauthorized modifications or incomplete disclosures.



Example in Practice:

A regulated healthcare company uses signed model cards to prove that an AI diagnostic tool was trained on certified datasets and validated for specific conditions, reducing liability risk.

Content Authenticity and C2PA Integration: Verifying Synthetic Media

The Problem:

As generative models proliferate, the risk of fabricated documents, images, and videos grows exponentially.

The Capability:

C2PA defines standards for embedding verifiable claims about content provenance. Integrating C2PA into trust workflows enables:

- Binding cryptographic signatures to generated outputs
- Embedding metadata on who created or modified content
- Providing proof for regulators, courts, and counterparties

Example in Practice:

A media organization uses DigiCert-issued C2PA signatures to label AI-generated images, ensuring consumers can verify authenticity instantly.

Agent Credentialing and Lifecycle Management: Controlling Autonomous Systems

The Problem:

AI agents often act independently—without clear identity or revocation controls.

The Capability:

Just as humans and devices are issued credentials, AI agents require:

- Scope-limited PKI certificates
- Automated renewal and revocation workflows
- Clear mapping to business policy controls

Example in Practice:

An enterprise deploys autonomous procurement bots that must present valid, signed credentials before executing transactions.

AI Assist: Scaling Trust Operations

The Problem:

Manual validation of signatures, credentials, and metadata doesn't scale with AI adoption.

The Capability:

- AI Assist is a generative, trust-focused copilot that can:
- Summarize SBOMs, AI BOMs, and C2PA manifests
- Answer compliance questions ("Is this model approved for production?")
- Proactively recommend next actions
- Maintain an audit trail of decisions

Why This Matters:

By combining AI-driven intelligence with cryptographic evidence, organizations can improve speed and accuracy without sacrificing assurance.

These capabilities don't operate in isolation. They're part of a broader regulatory and operational landscape that's evolving rapidly.

To build durable trust, organizations must align these technical foundations with emerging standards and regulatory expectations.

Navigating the New Rules of Trust

AI doesn't operate in a vacuum. Governments and industry bodies are racing to codify how organizations must govern and prove the integrity of their AI systems. Understanding and aligning to these frameworks is no longer optional—it's a precondition for market access, customer confidence, and legal defensibility.

Key Standards Shaping the Landscape

EU AI Act: The world's first comprehensive AI regulation, creating obligations for transparency, documentation, and risk management. High-risk AI systems—including many used in healthcare, finance, and infrastructure—must produce auditable records of training data, performance metrics, and provenance.

What this means: Enterprises need a way to produce signed Model Cards and AI BOMs as regulatory evidence.

NIST AI Risk Management Framework: A voluntary but influential framework defining principles for trustworthy AI: validity, reliability, safety, security, resilience, accountability, transparency, explainability, fairness, and privacy.

What this means: Compliance isn't simply about technical controls—it's about clear, provable governance practices.

C2PA (Coalition for Content Provenance and Authenticity): A cross-industry effort to standardize how digital content is signed and verified. As generative AI proliferates, C2PA will be the backbone of proving which content is authentic.

What this means: Organizations producing or distributing media must integrate signing workflows at the source.

ISO/IEC 42001: A management system standard for AI, specifying requirements for establishing policies, objectives, and processes to ensure trustworthy AI development and use.

What this means: Enterprises should prepare to integrate AI lifecycle governance into their existing compliance programs.

Strategic Implication: Regulatory scrutiny is converging on a single theme: If you can't prove it, you can't trust it. The cryptographic and lifecycle assurance frameworks that underlie DigiCert's core services are uniquely suited to this environment.

AI Trust in Practice

While many enterprises are still evaluating their AI strategies, common patterns are already emerging.
The examples below illustrate how the capabilities described in this paper could address real operational challenges.

Scenario 1 – Financial Services: Verifying Model Integrity Before Deployment

A global bank plans to integrate a third-party credit scoring model sourced from a reputable AI vendor. The model is highly accurate but lacks cryptographic signing or a verifiable record of training data. Without a trusted AI bill of materials and signed provenance, the bank faces regulatory exposure and increased fraud risk.

How Trust Infrastructure Helps:

- The model is signed with DigiCert-issued credentials, binding it to a unique version and configuration.
- A cryptographically signed AI BOM documents training data lineage and licensing constraints.
- Compliance teams can demonstrate to regulators that the model meets risk management requirements.

Scenario 2 – Manufacturing: Controlling Autonomous Agents on the Factory Floor

A multinational manufacturer deploys autonomous agents to optimize supply chain operations. These agents execute purchasing and inventory transactions with minimal human intervention. Currently, no clear credential lifecycle exists for these agents, creating risk if any of them are compromised.

How Trust Infrastructure Helps:

- Each agent receives a scope-limited PKI certificate that defines what actions it's authorized to perform.
- Credentials are automatically rotated and revoked if anomalies are detected.
- A trust policy engine logs every transaction with cryptographic proofs of agent identity.

Scenario 3 – Media & Entertainment: Proving Content Authenticity

A news organization adopts generative AI tools to produce real-time reporting visuals and data visualizations. As synthetic media grows harder to distinguish from deepfakes, the organization must prove the authenticity of its content to audiences and regulators.

How Trust Infrastructure Helps:

- Content is automatically signed and embedded with C2PA-aligned provenance metadata.
- Viewers can instantly validate the source and modification history.
- Legal teams can rely on tamper-evident records in any dispute.

Scenario 4 – Healthcare: Maintaining Compliance for Clinical AI Tools

A healthcare provider deploys an AI diagnostic tool to support radiology teams. Regulatory frameworks require transparent documentation of model training, validation, and updates.

How Trust Infrastructure Helps:

- Model Card Provenance is cryptographically signed and versioned.
- All updates are logged in an immutable ledger with timestamps and digital signatures.
- AI Assist capabilities generate compliance reports on demand.

Why These Scenarios Matter:

While hypothetical, each reflects real pressures facing enterprises today:
regulatory compliance, operational risk, and the need for demonstrable trust.
As AI adoption accelerates, these use cases will become routine—and organizations must be prepared to meet them with confidence.

DigiCert's Commitment

From foundational cryptography to modern supply chain security, DigiCert has been at the forefront of establishing verifiable trust online. We believe the next decade requires applying that same rigor to AI systems.

Our investments include:

Extending Signing Infrastructure: We're adapting our proven software signing capabilities to support AI model artifacts, ensuring model integrity from development through deployment.

Model Card Provenance and AI BOMs: We're defining standards and workflows for producing cryptographically signed model documentation, helping customers comply with regulatory mandates.

C2PA Integration: DigiCert is developing capabilities to issue and validate C2PA-aligned content credentials, enabling customers to embed verifiable provenance into digital media from the moment it's created.

AI Assist Development: We're designing AI-powered assistants to surface insights, guide policy adherence, and automate validation workflows inside DigiCert ONE—making trust operations more accessible, efficient, and reliable.

Industry Collaboration: DigiCert is engaging with regulators and standards bodies to help shape a global consensus on AI trust practices to prepare customers as regulations mature.

Why DigiCert?

Proven Scale:
Billions of certificates issued and validated globally.

Deep Expertise:
Two decades securing critical digital infrastructure.

Neutral Authority:
Trusted by enterprises, governments, and ecosystems.

Integrated Platform:
A single foundation to manage identities, signing, and trust evidence across traditional and AI-powered systems.

This isn't a pivot or a marketing experiment. It's a natural extension of our mission: **to make digital trust verifiable, durable, and universal.**

How to Get Started

Building an AI trust framework is a journey. No organization will solve every problem overnight, but taking deliberate steps today creates a foundation that can adapt over time.

6 Steps to Prepare:

1 Inventory Your AI Assets

- Document all AI models, data pipelines, autonomous agents, and generative tools in your environment.
- Identify which are mission-critical or subject to regulatory scrutiny.

2 Map Your Trust Gaps

- Assess which models have signed provenance and which rely on implicit trust.
- Determine whether content authenticity workflows exist.
- Evaluate the maturity of credential management for agents and models.

3 Prioritize Risk-Based Initiatives

- Triage efforts based on potential impact, regulatory exposure, and business value.
- Focus on high-risk areas first (e.g., models used in regulated sectors or customer-facing content).

4 Pilot Model Signing and AI BOM Workflows

- Work with your security and DevOps teams to integrate signing processes into existing pipelines.
- Establish procedures for generating and distributing AI BOMs.

5 Evaluate AI Assist Capabilities

- Identify operational workflows where AI-powered guidance and validation can improve efficiency.
- Test prototypes to measure time savings and accuracy improvements.

6 Align with Emerging Standards

- Develop a roadmap to integrate C2PA, NIST RMF, and EU AI Act requirements into your compliance programs.
- Monitor changes and adapt policies proactively.

How DigiCert Can Support You

AI Trust Readiness Workshops

Interactive sessions to assess gaps, map priorities, and define actionable next steps.

Proof-of-Concept Deployments

Pilot signing, provenance, and content validation workflows in a low-risk environment.

Early Access Programs

Be among the first to explore AI Assist capabilities in DigiCert ONE.

Remember: The sooner your organization establishes these practices, the easier it will be to adapt as AI grows more central—and as regulators demand more evidence.

Our Vision:

At DigiCert, we envision a world where every AI model is signed and traceable. Where every autonomous agent operates with verifiable credentials.

Where every piece of content can be proven authentic—instantly, and without doubt. Where trust itself evolves to keep pace with innovation.

Together, we can define the next decade of verifiable confidence.

We're building the infrastructure to make that vision real. Not someday—today.

If you're ready to explore how AI-powered trust can protect your business and unlock new possibilities, contact us at sales@digicert.com.

Conclusion

Imagine a world where every AI decision is backed by cryptographic proof. Where customers, regulators, and partners can validate the authenticity of every model, every action, every result.

It's not just possible—it's inevitable.

The future of digital trust is no longer about verifying static code or authenticating human identities. It's about proving the integrity of intelligent systems that learn, adapt, and act on your behalf.

AI introduces an unprecedented opportunity to unlock growth, but it also challenges every assumption we've made about how trust is established. The stakes have never been higher: Regulatory penalties, operational failures, and reputational damage await those who treat AI as a black box.

That means now isn't a time for hesitation—more than ever, it's a time for leadership.