WP-018 | Cryptographic AI-BOM | June 2026

Cryptographic AI-BOM: Zero-Knowledge Provenance Attestation for AI Model Supply Chains

AffixIO Research June 2026 WP-018 Download PDF

Regulators want an AI bill of materials. Vendors cannot publish training sets or architecture diagrams. Zero-knowledge attestations let you prove dataset hashes, fine-tune steps, and evaluation gates were satisfied while keeping trade secrets private.

CONTENTS

1. The AI Supply Chain Problem

Modern AI systems are not monolithic products. A production AI application typically comprises a foundation model sourced from one provider, fine-tuned on proprietary data by a second organisation, deployed through an inference API operated by a third, with retrieval augmentation from a fourth-party knowledge base, guardrails from a fifth, and evaluation tooling from a sixth. This layered supply chain creates a provenance problem of significant practical consequence.

When a high-risk AI system makes a consequential decision, regulators, auditors, and affected parties need answers to questions that the current AI supply chain cannot reliably answer: Which specific model version produced this output? Was the training data for that version licensed appropriately? Did the fine-tuning process introduce any known biases or behaviours? Was the model evaluated against the stated benchmark, and was that benchmark itself appropriate? Is the inference infrastructure the same as the one that was audited?

These questions are answerable in principle, but in practice they require trusting a chain of vendor attestations, each of which is typically provided as contractual documentation rather than as cryptographic evidence. A vendor can claim that their model was trained on licensed data; they cannot currently prove it to an external party without disclosing the training data itself or allowing the auditor unrestricted access to their infrastructure.

The AI-BOM concept, extending the software SBOM (Software Bill of Materials) to the specific components and processes of AI systems, is the emerging framework for addressing this problem. G7 governments and CISA formalised minimum requirements for AI-BOMs in May 2026. What remains missing is a cryptographic mechanism for producing AI-BOM entries that are verifiable by third parties without requiring proprietary disclosures.

---

2. From SBOM to AI-BOM: What Changes

Traditional SBOM tools, principally CycloneDX and SPDX, were designed for software supply chains: they catalogue libraries, frameworks, and binaries, each of which has a

deterministic identity that can be verified by comparing cryptographic hashes of the binary against a known-good reference.

AI supply chains have properties that break the assumptions underlying traditional SBOM approaches.

Model Weights Are Not Binaries

A trained model's weights are not analogous to a compiled binary. Identical training code run on identical data with identical hyperparameters will produce slightly different weights on different hardware due to floating-point non-determinism. There is no equivalent to the binary hash that uniquely identifies a specific model version without disclosing the weights themselves.

Training Data Is Not a Package

Software dependencies are discrete, versioned artefacts with stable identifiers. Training datasets are often assembled from multiple sources, preprocessed, filtered, and augmented in ways that are specific to the training run. Documenting training data provenance requires a richer schema than SBOM package references provide, and verifying that documentation requires either disclosing the data or trusting the provider's assertion.

The Inference Stack Is Dynamic

A software SBOM captures the dependencies at a point in time. An AI inference stack includes dynamic components: retrieval indexes that are updated continuously, prompt templates that change with model versions, guardrails that are updated in response to new failure modes, and quantisation configurations that vary by deployment environment. A static AI-BOM document is insufficient; provenance attestations must be generated continuously as the stack evolves.

---

3. The G7 and CISA AI SBOM Frameworks

The G7 AI SBOM framework released in May 2026 organises AI-BOM entries into seven core clusters, each addressing a different dimension of AI supply chain transparency.

G7 Cluster	Required Elements	Cryptographic Challenge
Metadata	System identifier, version, purpose, deployment context	Low: straightforward to sign and verify
Models	Architecture, weights identifier, training lineage, licence, provenance	High: weights disclosure risk; lineage may be proprietary
Dataset Properties	Source, collection method, licence, preprocessing, splits, known biases	High: dataset disclosure risk; bias characterisation is sensitive
System Level Properties	Inference configuration, hardware, quantisation, serving stack	Medium: configuration may reveal commercial details
Key Performance Indicators	Accuracy, latency, throughput, fairness metrics	Low-Medium: benchmark attestation possible with ZKML
Security Properties	Adversarial robustness, red team results, vulnerability assessments	Medium: results may reveal exploitable weaknesses
Infrastructure	Cloud provider, region, access controls, audit logging	Low-Medium: infrastructure details may be commercially sensitive

CISA's minimum elements guidance complements the G7 framework with a focus on supply chain risk: it requires that AI-BOMs include provenance attestations for every component that forms part of the AI system's critical decision path, and that these attestations be updatable as components change.

4. EU AI Act Article 11 Technical Documentation

EU AI Act Article 11 requires providers of high-risk AI systems to maintain technical documentation that enables authorities to assess compliance. Annex IV specifies the required documentation elements, which substantially overlap with the G7 AI-BOM framework. Key elements include a description of the general logic of the AI system and of the design choices made, the training methodologies, techniques, and data used, the testing procedures and testing results, and information on the data governance and data management practices.

Article 11 documentation must be maintained for the entire lifetime of the AI system and must be updated when the system is modified in ways that affect its compliance. This creates a continuous documentation obligation that is impractical to fulfil with static documents: every model update, dataset refresh, or inference stack change potentially triggers a documentation update requirement.

AffixIO's Cryptographic AI-BOM addresses this by generating provenance attestations automatically at each supply chain event, maintaining a cryptographic record of the AI system's component history that satisfies Article 11's documentation requirements without requiring manual documentation updates for routine operational changes.

---

5. Why Standard SBOM Tools Are Insufficient

CycloneDX's ML-BOM extension provides a structured schema for AI-BOM entries. It is a significant step forward in AI supply chain documentation, and AffixIO's architecture is designed to be compatible with CycloneDX ML-BOM output. However, standard SBOM tools have three limitations that a cryptographic AI-BOM must address.

Attestation Without Verification

A CycloneDX ML-BOM document is a structured assertion: the model provider states that their model has property X. The document may be signed by the provider, but the signature only establishes who made the assertion, not whether the assertion is true. A provider could sign a document claiming their model was trained on licensed data when it was not; the signature provides no protection against false assertions.

No Support for Confidential Properties

Many AI-BOM properties are commercially sensitive. A provider may be willing to attest that their training data meets a specified bias threshold, but not to disclose the training data itself or the detailed results of bias testing. Standard SBOM tools have no mechanism for attesting to properties of undisclosed components: you either disclose the component or you omit the attestation.

Static Point-in-Time Snapshots

SBOM documents are typically generated at release time and updated manually when components change. For AI systems with continuously updated retrieval indexes, frequently refreshed model versions, or dynamically updated guardrails, a static SBOM is outdated within hours of generation. A cryptographic AI-BOM requires a continuous attestation architecture, not a point-in-time document.

---

6. Cryptographic Provenance: The ZK Approach

AffixIO's Cryptographic AI-BOM replaces assertion-based documentation with proof-based attestation. For each AI supply chain component, the responsible party generates a ZK proof that attests to properties of that component without disclosing the component itself.

> **The Attestation Principle**
>
> Instead of "we attest that our training data is licensed and bias-tested" (assertion), the provider generates: "this training dataset has a commitment C, the dataset satisfies the GDPR Article 5 minimisation predicate, the bias score across 12 protected categories is below threshold T, and all sources are in the approved licence registry" (proof). The verifier checks the proof without receiving the dataset.

Commitment-Based Component Identity

Each AI supply chain component, whether a model checkpoint, a training dataset shard, a fine-tuning configuration, or an inference stack specification, is committed using a collision-resistant hash. The commitment serves as the component's verifiable identifier throughout the AI-BOM. When a component is updated, its commitment changes, creating an auditable history of component versions.

SLSA Attestation Integration

Supply Chain Levels for Software Artifacts (SLSA) is the standard framework for software provenance attestation. AffixIO's AI-BOM architecture generates SLSA provenance statements for AI components, extending SLSA's build provenance model to cover training runs, evaluation pipelines, and deployment events. Each SLSA attestation is signed and linked to the relevant AI-BOM commitment, enabling a verifier to trace any AI system component back to its origin.

---

7. AffixIO AI-BOM Architecture

The AffixIO Cryptographic AI-BOM architecture comprises four layers: component commitment, event attestation, AI-BOM root construction, and audit interface.

Layer 1: Component Commitment

Each AI supply chain component is committed at creation or acquisition. Model weights are committed using a SHA-256 hash of the serialised weight tensor, generating a component identifier that is stable and unique. Training datasets are committed using a Merkle tree over the dataset shards, enabling inclusion proofs for specific data points without disclosing the full dataset. Inference stack configurations are committed as structured JSON hashes.

Layer 2: Event Attestation

Every supply chain event, training run initiation, checkpoint generation, fine-tuning, evaluation, deployment, and update, generates a ZK attestation. The attestation proves that the event was performed with the committed components, under the specified

configuration, and that the output satisfies stated quality and compliance predicates. Attestations are signed by the responsible party's HSM-held key.

Layer 3: AI-BOM Root Construction

All component commitments and event attestations for a specific AI system version are combined into an AI-BOM Merkle tree. The root of this tree is the AI-BOM root: a single 32-byte value that commits to the entire provenance history of the AI system. The AI-BOM root is included in the system's deployment record and anchored to an append-only log, enabling retrospective verification of the provenance record as of any deployment timestamp.

Layer 4: Audit Interface

Auditors and regulators interact with the AI-BOM through a query interface that accepts natural-language or structured queries and returns relevant attestations together with inclusion proofs from the AI-BOM tree. A regulator asking "which training datasets were used in the current production model?" receives the dataset commitments, the licence attestations, and Merkle inclusion proofs, without receiving the datasets themselves.

---

8. Model Weight Provenance Without Weight Disclosure

The single most commercially sensitive element of an AI-BOM is the model weights. Weights represent the concentrated value of the training investment; disclosing them is equivalent to disclosing the product itself. Yet regulators and customers legitimately need to verify properties of the weights, such as whether they are the approved production version, whether they have been modified since evaluation, and whether they match the weights used in the cited benchmark results.

AffixIO's weight provenance circuit addresses this directly. At model deployment, the deploying organisation generates a ZK proof asserting: the deployed weights have commitment W, W is identical to the commitment recorded in the evaluation attestation, and W differs from the pre-training checkpoint by no more than the delta

expected from the fine-tuning process. This proof enables a verifier to confirm that the production model is the evaluated model, without receiving either the weights or the delta.

For foundation model providers who supply models to multiple downstream fine-tuners, AffixIO provides a model provenance certificate: a ZK proof that a specific fine-tuned model descends from a specific foundation model version, without revealing either the foundation model weights or the fine-tuning delta. This addresses the increasingly common scenario where an enterprise customer needs to verify their AI vendor's claims about model origin and modification history.

---

9. Training Data Lineage Attestation

Training data lineage is the most complex element of AI supply chain provenance. A foundation model may be trained on web-scraped data, licensed corpora, synthetic data, and human feedback signals, each with different licensing terms, different data governance requirements, and different bias characteristics. Documenting and verifying this lineage without disclosing the datasets requires a structured attestation approach.

AffixIO's training data attestation circuit takes as private inputs the Merkle tree of the training corpus and the licence registry for each data source. It produces a proof asserting: every shard in the corpus has a licence commitment that maps to an approved licence type, the proportion of each licence category is within the stated bounds, and the bias score across the specified protected categories is below the stated threshold. The verifier receives the proof and the stated bounds; they learn whether the training data meets the requirements without learning what the training data contains.

For regulated datasets, such as medical records used to train clinical AI systems, the attestation circuit additionally proves that the data handling process complied with the relevant data protection framework, such as GDPR Article 9 for special category health data or HIPAA for US health information. This compliance proof can be

presented to regulators as evidence of lawful processing without disclosing patient records or data processing details.

---

10. Fine-Tuning and Inference Stack Provenance

The provenance chain does not end at training. Fine-tuning, alignment training, quantisation, and inference stack configuration each introduce changes to the AI system's behaviour that must be documented in the AI-BOM.

Fine-Tuning Attestation

Each fine-tuning run generates an attestation recording the base model commitment, the fine-tuning dataset commitment, the training configuration, and the output model commitment. The attestation proves that the fine-tuned model was derived from the base model using the specified dataset and configuration, enabling a verifier to trace the model's lineage without receiving any of the underlying artefacts.

RLHF and Alignment Attestation

Reinforcement Learning from Human Feedback (RLHF) and other alignment training processes introduce human judgements into the model's training signal. The AffixIO alignment attestation circuit proves that the alignment training used a reward model trained on human feedback that met specified quality criteria, without disclosing the specific feedback examples or the reward model architecture.

Inference Stack Versioning

Prompt templates, system prompts, retrieval configurations, and guardrail specifications are all components of the effective AI system that influence outputs. AffixIO's inference stack provenance module maintains a continuous record of these configurations, generating an attestation at each configuration change. This ensures

that the AI-BOM accurately represents the full inference-time system, not merely the trained model.

---

11. Procurement and Audit Integration

AIBOM is moving from optional security artefact to enforceable procurement requirement in 2026. Several large enterprises and public sector buyers have announced requirements for AI-BOM documentation as a condition of AI system procurement. AffixIO's Cryptographic AI-BOM is designed to satisfy these procurement requirements while minimising the disclosure burden on AI providers.

Procurement Verification Workflow

In a procurement context, the AI provider generates an AI-BOM root for the proposed system and submits it to the buyer alongside the relevant provenance attestations. The buyer's procurement team verifies the attestations against stated requirements, for example confirming that training data licences are acceptable and that security properties meet the specified standard, without requiring the provider to disclose the underlying components. The AI-BOM root is included in the contract as a verifiable reference to the delivered system.

Regulatory Audit Workflow

For regulatory audits, AffixIO's audit interface enables regulators to submit structured queries about the AI system's provenance and receive proofs in response. The regulatory audit does not require regulator access to the AI system's infrastructure, the model weights, or the training data: it requires only access to the attestation query interface and the verification keys for the relevant circuits. This is a more efficient and more reliable audit mechanism than infrastructure access, which creates both security risks and operational disruption.

---

12. Conclusions

The AI supply chain is the next frontier of enterprise security and regulatory risk. G7 governments, CISA, and the EU AI Act have all identified AI-BOM documentation as a critical requirement for AI governance, but the current state of AI-BOM tooling relies on assertion-based documentation rather than cryptographic verification. This gap between the stated requirement and the available mechanism creates risk for both AI providers and AI buyers.

AffixIO's Cryptographic AI-BOM closes this gap by applying zero-knowledge proof technology to the specific attestation challenges of the AI supply chain: model weight provenance without weight disclosure, training data lineage attestation without dataset disclosure, fine-tuning chain verification without model architecture disclosure, and continuous inference stack provenance without configuration disclosure. The result is an AI-BOM that satisfies G7, CISA, and EU AI Act Article 11 requirements while protecting the commercial confidentiality of all supply chain participants.

The Cryptographic AI-BOM architecture is available as an open-source component compatible with CycloneDX ML-BOM, SLSA provenance attestation, and the OWASP AIBOM specification, with enterprise support for integration into existing AI governance and procurement workflows.

Related reading

- WP-010: The Open-Source AI Governance Stack
- WP-001: Cryptographic AI Governance: A Technical Framework
- WP-021: Beyond C2PA: ZK Content Provenance That Survives Metadata Stripping

Frequently asked questions

What is an AI-BOM?

An AI bill of materials lists datasets, base models, fine-tuning runs, evaluation results, and deployment artefacts that produced a given model version.

Can ZK proofs hide proprietary training data?

Yes. You prove commitments to dataset hashes and licensing checks without revealing raw records or model weights.

Does this align with EU AI Act Article 11?

Cryptographic attestations support technical documentation requirements by providing verifiable supply chain records alongside human-readable docs.
