

NAGE

WHITE PAPER

Source-aware AI and verifiable records



The Nage architecture and its current verification boundaries

KLM / SEDIM / Nage AI OS / Craton

September 2026 | Conference edition

Executive perspective

Nage develops AI infrastructure that makes an output's declared basis available for inspection.

The problem

A useful answer and an inspectable answer are different properties. Enterprise workflows need the identity and version of evidence, a record of the system configuration and an explicit account of what each signal measures. A citation alone does not establish that a passage supports a claim.

The architecture

Nage develops KLM as a candidate standard and operates the reference implementation. SEDIM is its architecture; Nage AI OS carries the resulting records across services. This is the trust-anchor proposition: define the shared record and make it available for checks outside the issuing service. Craton and Craton Pro are the base-model research programme.

The current boundary

The current service uses Qwen3-8B as its base. The programme reports signed records, source hashing and structured signals. It also reports incomplete historical source links, an unpublished Merkle commitment and a trained weight-injection path that is not yet the production path. These boundaries determine how the system should be evaluated.

The central claim

A reviewer can check selected record properties without operating Nage's service. This is narrower than proving that the declared inference occurred, proving causal attribution for every claim or establishing the truth of the answer.

This paper describes the September 2026 implementation position and research programme. Experimental measurements are identified as Nage-reported results.

The enterprise problem

An output becomes easier to review when the information behind it is addressable.

Source identity

A reviewer needs to distinguish an authoritative internal record from a user assertion, volatile external content, a tool response or model-generated text. These inputs can coexist in one answer but do not carry the same evidential status.

Change over time

A source may be updated, superseded or withdrawn. A record should identify the relevant version and distinguish the evidence retained at the time of a response from the evidence available today.

Support and uncertainty

Retrieval similarity measures relevance to a query. Claim support asks a separate question about the relation between evidence and an assertion. A computed number also needs a method, scope and status before it can inform a review decision.

Review and action

The useful outcome is a workflow in which missing evidence, conflicting sources and unavailable measurements lead to explicit handling. A record can support that workflow; the institution still defines its decision rights and escalation policy.

KLM as a common contract

Knowledge Layers Model separates the format of a record from the private design of an implementation.

What the contract covers

The supplied internal specification describes claims, evidence, source classes, traces and verification requirements. The public repository is at charter stage; it does not yet publish the full v1.0 specification. The levels below describe the supplied internal document.

Conformance is cumulative

The six levels below describe the specification’s progression. A signature alone does not demonstrate compliance with every preceding requirement. A declaration should name the specification version, assessed scope and evidence supporting the level.

KLM-0 / Declared	Records and explicit unknowns Emit the required structure and epistemic statuses without fabricating unavailable signals.
KLM-1 / Traceable	Inputs and events Record model versions, source identities, retrieval traces and relevant output spans.
KLM-2 / Grounded	Evidence and execution links Establish claim-to-evidence and procedure-to-execution links with the required signals.
KLM-3 / Reflective	Confidence and governance traces Keep declared and grounded confidence distinct and preserve policy provenance.
KLM-4 / Attributable	Traceable causes Link reflective findings to their generative basis and record governed execution attribution.
KLM-5 / Verifiable	Independent checking Enable third-party checking of records within the declared scope.

Sources: supplied KLM Specification §8.3–8.4; public charter at github.com/NageAI/klm. These levels are not external certification.

SEDIM and its operating paths

SEDIM means Sedimentary Intelligence Model. Its target architecture separates a frozen base from managed knowledge layers.

Production and development

The conference materials report that trained CLAST weights are not yet applied to live inference. The current grounding and record path must therefore be distinguished from the target weight-level composition mechanism. Logging source participation does not by itself establish the causal contribution of each source to each output claim.

FACIES

Base-model slot

Holds the frozen base. The current deployment uses Qwen3-8B. Craton is designed to occupy this role in a future qualified deployment.

VARVE

Managed knowledge layer

An independently managed source or domain layer with identity and lifecycle information.

CLAST

Layer initialisation

Cross-Layer Asymmetric Spectral Transfer initialises trainable layers from residual spectral structure.

STRIA

Routing

Selects or weights knowledge-layer participation for a query.

STEMMA

Attribution record

Represents the declared source participation associated with an output.

CENTO

Composite output

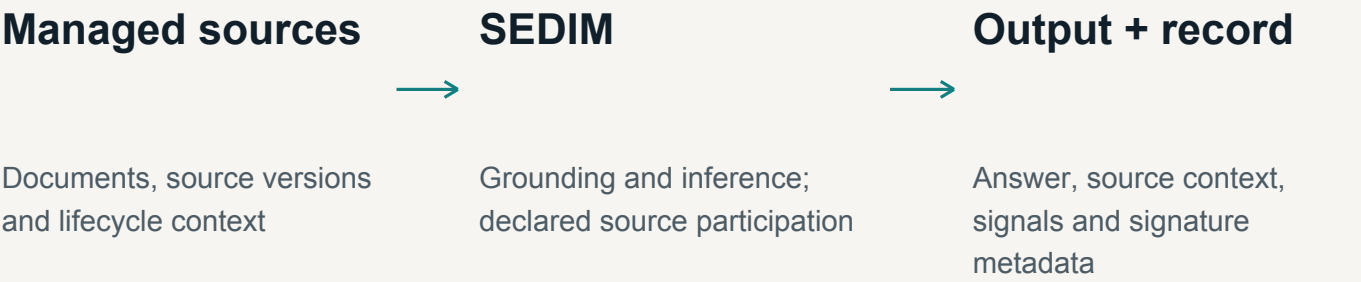
Combines the answer with its associated provenance and epistemic information.

Source: September 2026 conference material, SEDIM architecture and implementation-scope notes.

The Nage architecture

KLM defines the record contract; Nage operates the reference implementation.

KLM / CANDIDATE STANDARD
Shared record vocabulary, signal status and conformance requirements



Nage AI OS carries the record across services
Application / reviewer / assurance tooling

Current base Qwen3-8B in the FACIES slot. Current grounding and record service.	Development programme Craton / Craton Pro and trained VARVE weight composition.
---	---

Conceptual service flow, not a causal execution proof. Current inference uses grounding with Qwen3-8B; trained VARVE weight injection and Craton remain development paths.

Signals with explicit meaning

The record separates the system’s declared confidence from the evidence-related signals used to assess it.

A confidence gap needs context

Where an implementation reports a numeric difference between declared and grounded confidence, it should preserve the component vector, formula and calibration identity. The difference is an operational signal, not automatically a calibrated probability of error.

Measurement is not validation

The current lexical claim-support method failed a constructed contradiction evaluation. A separate contradiction detector has been tested but is not yet connected to the live gamma record. Its result must not be presented as a deployed correction.

Retrieval relevance

Query-to-source relation

How relevant the retrieved passage is to the query under the declared method.

Claim support

Claim-to-evidence relation

An assessment of whether the cited material supports a claim. Its method and demonstrated limitations matter.

Evidence coverage

Coverage of output claims

How many identified claims have associated evidence under the implementation’s definition.

Source authority

Status of the source

An assessment based on source class or policy. It may be heuristic rather than measured.

Source freshness

Temporal information

Recency under a defined policy. Missing dates should remain unavailable with a reason.

Source: gamma presentation block and claim-support evaluation. The technical annex provides the reported sample sizes and error rates.

An output with a reviewable record

A record keeps source context, signal meaning and verification scope together.

SYNTHETIC EXAMPLE / TEST KEY

Declared confidence	Grounded confidence	Confidence gap
0.910	0.742	+0.168

declared - grounded = gap

Retrieval relevance	0.9422	measured
Source authority	0.600	heuristic
Source freshness	null	unavailable: no source date

What is checked

Record integrity, key attribution and specified internal relationships.

What remains separate

Answer truth, causal execution evidence and method validation.

Illustrative view of the included synthetic fixture. Values are demonstrations, not measured production quality or calibrated probabilities.
Source: verification/sample/receipt.json.

What a verifier establishes

Cryptographic checks, log inclusion and execution evidence answer different questions.

Current implementation boundary

Nage reports Ed25519 record signing and Python/TypeScript reference verifiers. The record chain uses HMAC and is not publicly verifiable with a public key. The Merkle root described in the September materials is unsigned and unpublished. No independent external audit is reported.

Signed record	Integrity and key attribution Given a trusted public key, signature verification checks that the signed body has not changed and was signed by the corresponding private key.
Record consistency	Declared relationships A verifier can check specified arithmetic, required null reasons and the presence or consistency of manifest fields.
Conditional recomputation	Selected signal values Recomputation needs the disclosed inputs, method and required artefacts. Repeating a score does not validate the score’s semantic meaning.
Log inclusion	Membership in a committed set A Merkle proof can show that a record belongs to a particular tree. Append-only consistency and omission detection require additional protocol evidence.
Execution evidence	What actually ran A signed description or Merkle inclusion proof does not alone prove that the declared model and sources produced the output.

Sources: Nage verification annex; RFC 6962, Merkle audit and consistency proofs. See the reference register.

Source provenance and lifecycle

Provenance begins with the material that enters the system and continues through its transformations.

Three identities

The intended ingestion chain identifies uploaded bytes, extracted text and stored chunks separately. This prevents the hash of one representation from being mistaken for the identity of another. Rechecking an extraction also needs its source, method and version.

Historical coverage

Nage reports hashes for 7,230 stored chunks with zero mismatches against a database digest. Those hashes are marked backfill: they commit to the text held when the hashes were computed. The historical corpus still has gaps in source and extraction hashes and links to original files. New-ingestion checks must not be generalized to that older corpus.

Removal scope

The reported privacy-delete behaviour zeroes or removes a layer's contribution in the serving path and maintains a consolidation ledger. Durable weight artefacts in object storage are not purged by the same operation. Removing one managed layer also does not establish that equivalent information is absent from the base or other sources.

A deployment policy

An enterprise evaluation should define source ownership, version retention, withdrawal handling, artefact deletion and the evidence retained after deletion. Record retention and data removal should be designed together, with the institution's applicable obligations in view.

Nage AI OS

The operating environment carries records and epistemic status across service boundaries.

Integration principle

A downstream service should preserve source classes, record identifiers and signal status. An unavailable measurement must not become a confident score simply because a record crossed an API boundary.

Readiness is capability-specific

The conference map identifies eight components as live and MAGMA as designed. That does not imply that every planned feature of the live components is deployed. Production CONDUIT signing, for example, remains an identified integration task.

Institutional deployment

Identity, authorisation, disclosure, logging and human review belong in the deployment design. SEDIM records can feed an organisation’s existing assurance process without replacing its governance responsibilities.

Knowledge	SEDIM / CONFLUX / BEDROCK Knowledge handling, training or transfer tooling, and dataset management.
Substrate	NUCLEUS / RILL State and semantic services, together with communication between components.
Delivery and observation	STRATUM / CONDUIT / OUTCROP Disclosure and access boundaries, external-system integration and developer visibility.
Execution contract	MAGMA A service contract and reference stub. The production-serving runtime is not yet MAGMA.

Craton and Craton Pro

A base-model programme for reducing the knowledge that sits outside managed layers.

The attribution gap

An existing pretrained model can supply information that was not introduced through a managed VARVE. Source-layer routing cannot fully explain that contribution. The practical scope of attribution depends partly on what the base already knows.

The programme

Craton is a from-scratch base designed to constrain factual content and hold domain knowledge in addressable layers. Craton Pro is the larger class of the same architecture. These are development-stage models; the conference material does not establish production readiness or a general zero-attribution-gap guarantee.

What the experiments changed

The held-out CLAST weight-path test scored 0/15. Separately, an inspection of the S3 retrieval-and-context path found that packing removed evidence before it reached the prompt. That finding affects interpretation of the S3 context comparison. It does not explain away or overturn the separate 0/15 weight-path result.

Acceptance before scale

The next decision should depend on demonstrated evidence use, source isolation and retained utility under defined tests. Ten successful sterility probes are a limited observation. They do not establish the absence of all factual knowledge or generalise to every source family.

The technical annex records the reported results and the evidence-packing correction. No scale, release date or performance commitment is made here.

Evaluation and adoption

A useful evaluation begins with a narrow workflow and a shared definition of success.

Define the perimeter

Select a workflow, a source collection and a reviewer. Identify which output claims matter, what evidence the reviewer can access and which decisions remain with people.

Test the record

Check a real exported record using the published production key. Deliberately alter signed content and confirm rejection. Test the applicable disclosure view and confirm which fields can be recomputed.

Test the workflow

Use examples with supported, contradicted and absent evidence. Include source updates and withdrawals. Measure reviewer usefulness as well as task quality, latency and operational cost under the intended load.

Agree the deployment boundary

Document supported integrations, retention behaviour, access controls and remaining development work. Any service levels or commercial terms should be established for that deployment rather than inferred from a conference benchmark.

Regulatory context

Record keeping and transparency can support institutional assurance. EU AI Act Articles 12 and 13 concern record keeping and information to deployers, respectively. Cryptographic verification is not a certification of legal compliance, and the Act does not prescribe SEDIM's architecture.

Discuss an evaluation with the Nage team via the accompanying invitation or nage.ai.

References and scope

The specification, implementation reports and external technical precedents serve different purposes.

Nage specification and implementation sources

[N1] Supplied internal KLM Specification, §8.3–8.4 and §12, September 2026. Public source: Nage AI, KLM Charter (Apache-2.0), <https://github.com/NageAI/klm>. Full public v1.0 text remains forthcoming.

[N2] Nage company definition and ecosystem map, September 2026.

[N3] Conference verification annex and measured-results annex, 14 September 2026.

[N4] Claim-support evaluation and Craton sealed-prediction methodology, 14 September 2026.

External references

[R1] Laurie, Langley and Kasper. Certificate Transparency. RFC 6962, 2013. Merkle inclusion and consistency mechanisms. Superseded by RFC 9162.
<https://www.rfc-editor.org/rfc/rfc6962>

[R2] Fleshman et al. AdapterSwap: Continuous Training of LLMs with Data Removal and Access-Control Guarantees. arXiv:2404.08417. Modular knowledge adaptation is relevant prior work.
<https://arxiv.org/abs/2404.08417>

[R3] European Commission AI Act Service Desk. Articles 12 and 13.
<https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-12>
<https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-13>

Reading the evidence

Nage-reported results describe particular artefacts, tests and implementation states. They should not be read as an independent external audit, a universal performance guarantee or legal approval. The accompanying technical annex preserves the material limits of each reported result.