AI-Restricted Open Standard (AIROS)

Technical Framework v1.2 — VisionZeroAI Whitepaper

AIROS defines a vendor-neutral, verifiable method to mark digital content as AI-restricted ('do-not-ingest'), ensuring that AI systems can automatically detect and respect privacy, ownership, and policy boundaries. This whitepaper, maintained by VisionZeroAI (Pty) Ltd, outlines the technical foundation, architecture, and governance of AIROS as a global open standard for responsible AI data protection.

Maintained by VisionZeroAI (Pty) Ltd — https://www.visionzero.co.za

Contact: info@visionzero.co.za

1. Executive Summary

AIROS (AI-Restricted Open Standard) defines a vendor-neutral, verifiable way to mark digital content as AI-restricted so that AI platforms, data pipelines, and compliance systems can automatically detect and respect ownership, privacy, and policy boundaries. This version adds a canonical tag schema, embedding methods, a security model, governance, and reference APIs to accelerate adoption.

2. Problem & Objectives

Problem. Sensitive content may be uploaded to AI systems or flow into model training without explicit consent or policy controls. There is no cross-vendor 'do-not-ingest' signal trusted by AI platforms.

Objectives:

- Provide a simple, interoperable tag that travels with the file or as a sidecar.

- Enable cryptographic verification without disclosing content.

- Offer a minimal, privacy-preserving trust registry for key discovery and revocation.

- Align with emerging AI governance, privacy, and security standards.

3. Design Principles

• Interoperability first • Privacy by design • Verifiability • Minimalism • Backward compatibility

4. Technical Architecture Overview

Core: Owner Tools (SDK/CLI); Verification SDK; Trust Registry; Optional Signing API; Enforcement Points.

Flow: Tag -> Verify -> Protect (hash, sign, verify, block/quarantine per policy).

5. Canonical Tag Schema (JSON)

```
{
  "airos_version": "1.2",
  "restricted": true,
  "owner_id": "VisionZeroAI",
  "file_hash": "sha256:ABCD...",
  "alg": "Ed25519",
  "kid": "vz-2025-q4-ed25519-01",
  "signature": "base64url(JWS_signature)",
  "issued_at": "2025-10-24T12:00:00Z",
  "policy_ref": "https://visionzero.co.za/airos/policies/default"
}
```

Notes: file_hash over exact bytes; alg Ed25519 or RSA-3072; kid resolvable via registry; signature is detached JWS over canonical tag.

6. Embedding Methods (Metadata & Sidecar)

PDF (XMP) snippet:

```
<x:xmpmeta xmlns:x="adobe:ns:meta/">
 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description xmlns:airos="https://visionzero.co.za/ns/airos/1.2/">
    <airos:restricted>true</airos:restricted>
    <airos:owner_id>VisionZeroAI</airos:owner_id>
    <airos:file_hash>sha256:ABCD...</airos:file_hash>
    <airos:kid>vz-2025-q4-ed25519-01</airos:kid>
    <airos:signature>...</airos:signature>
  </rdf:Description>
 </rdf:RDF>
</x:xmpmeta>
```

DOCX: store JSON in custom property AIROS; Images: XMP side-bag; Sidecar: {filename}.airos.json.

```
{
  "filename": "report.pdf",
  "tag": { ... canonical AIROS tag ... },
  "integrity": "sha256:ABCD..."
}
```

7. Security Model

Algorithms: Ed25519 (preferred) / RSA-3072; SHA-256. Keys in HSM/KMS; rotate; kid is stable identifier.

Verification: recompute file_hash, resolve public key via registry, check signature; fail on revoke/expire.

Privacy: never upload file bytes; verification can be offline (sidecar).

8. Reference APIs (v1)

Signing API:

```
POST /v1/sign
Headers: X-API-Key: <org_key>
Body: { "file_hash":"sha256:ABCD...", "kid":"vz-2025-q4-ed25519-01",
"policy_ref":"https://..." }
-> 200 { "signature":"...", "kid":"...", "issued_at":"..." }
```

Trust Registry (public):

```
GET /v1/orgs/{owner_id}/keys
GET /v1/keys/{kid}
GET /v1/crl
```

9. SDKs & References

airos-sdk-python; airos-sdk-node; CLI; middleware examples.

10. Compliance & Regulatory Alignment

Complements EU AI Act, GDPR, NIST AI RMF, ISO/IEC 42001 by providing a technical consent boundary signal.

11. Governance, Versioning & Participation

Open working group; semantic versioning; initial stewardship by VisionZeroAI; roadmap to multi-stakeholder governance.

12. Adoption Paths & Integration Patterns

AI platforms: upload/pre-ingestion verification; Enterprises: DMS/ECM templates; Vendors: DLP/CASB detectors.

13. Use Cases

Legal, Healthcare, Finance, Government examples to anchor practical adoption.

14. Pilot Program

Scope: SDK integration + ingestion gate; Metrics: detection rate, FP/FN, latency; Deliverables: playbook, corpus, middleware.

15. Roadmap

Q4 2025 v1.2; Q1 2026 SDKs & Registry beta; H1 2026 certification; H2 2026 v2.0 with analytics.

16. Call for Collaboration

Join via GitHub or email info@visionzero.co.za; contribute, pilot, or align as an adopter.

17. Glossary

AIROS Tag, Sidecar, KMS/HSM, kid, CRL.

Appendix A — Example Tags

```
{
  "filename": "brief.docx",
  "tag": {
    "airos_version": "1.2",
    "restricted": true,
    "owner_id": "AcmeLegal",
    "file_hash": "sha256:7C8B...",
    "alg": "Ed25519",
    "kid": "acme-2026-ed25519-01",
    "signature": "MEUCIQ...",
    "issued_at": "2026-01-15T09:00:00Z",
    "policy_ref": "https://acme.example/policy/airos"
  },
  "integrity": "sha256:7C8B..."
}
```

Appendix B — Verification Pseudocode

```
function verify(file):
  tag = extract_airos_tag(file)
  if not tag or tag.restricted != true: return {allow: true, reason: "no_tag"}
  bytes = read_file(file)
  if sha256(bytes) != tag.file_hash: return {allow: true, reason: "hash_mismatch"}
  pubkey = registry.get_key(tag.kid)
  if pubkey.status != "active": return {allow: true, reason: "key_invalid"}
  if verify_signature(pubkey, canonicalize(tag_without_signature), tag.signature):
      return {allow: false, reason: "AI-Restricted verified"}
  else:
      return {allow: true, reason: "bad_signature"}
```

Appendix C — HTTP Status & Error Model (Reference)

```
200 OK
400 Bad Request
401 Unauthorized
403 Forbidden
404 Not Found
409 Conflict
429 Too Many
5xx Server
```

Appendix D — Change Log

v1.2: canonical schema, embedding methods, security model, reference APIs, governance, compliance alignment, expanded use cases.