This page is authoritative in English. Translated summary available on request via contact.ethraeon.ai.

ETHRAEON ARCANUM

Humanitas ante Machinam

ETHRAEON TECHNICAL PAPER SE28

CoRIx Constitutional Measurement Framework

Hierarchical AI System Evaluation Implementing NIST ARIA 0.1 Specifications

June 2026 S. Jason Prohaska Patent Pending

ABSTRACT

This paper presents the Constitutional Risk Index (CoRIx) measurement framework, a hierarchical evaluation system designed to assess AI outputs against constitutional governance standards. CoRIx implements the National Institute of Standards and Technology (NIST) Artificial Intelligence Risk and Impact Assessment (ARIA) 0.1 specification while extending it with constitutional compliance verification unique to the ETHRAEON architecture. The framework establishes a six-level measurement tree that progresses from high-level interpretation through granular telemetry, enabling both human oversight and automated governance at scale. We demonstrate that CoRIx provides the first commercially available ARIA-compliant measurement

system, positioning constitutional runtime for federal procurement and enterprise deployment.

LAYER 1 - ONTOLOGY

CoRIx - Conceptual Foundation

The Constitutional Risk Index emerges from a fundamental insight: AI systems cannot be governed effectively through post-hoc evaluation alone. Meaningful governance requires measurement that is constitutional by design - embedded in the evaluation architecture itself rather than applied as an afterthought.

1.1 The Measurement Problem

Contemporary AI evaluation suffers from three critical deficiencies:

- **Fragmentation:** Metrics scattered across isolated tools without unified governance framework
- **Opacity:** Black-box scoring without hierarchical decomposition enabling human understanding
- **Compliance Gap:** No commercial system implementing federal AI assessment standards

CoRIx addresses each deficiency through constitutional architecture. Rather than evaluating AI outputs against arbitrary thresholds, CoRIx measures against codified principles - the ΔSUM Codex invariants that define acceptable AI behavior within the ETHRAEON framework.

1.2 NIST ARIA Alignment

The NIST ARIA 0.1 framework establishes federal standards for AI risk assessment. CoRIx implements ARIA's hierarchical evaluation model while extending it with constitutional

compliance verification:

ARIA SPECIFICATION COMPLIANCE

CoRIx implements 100% of NIST ARIA 0.1 required elements including six-level hierarchical structure, five-category automated annotation (Response Quality, Appropriateness, Factuality, Safety, User Experience), and validity scoring methodology. ETHRAEON extensions include constitutional compliance verification, ΔSUM trace provenance, and Trinity orchestration layer mapping.

1.3 Philosophical Foundation

CoRIx operates under the principle that measurement itself is a form of governance. By structuring what we measure and how we measure it, we shape the boundaries of acceptable AI behavior. This is not neutral observation - it is constitutional constraint expressed through evaluation architecture.

The framework embodies *Humanitas ante Machinam*: human judgment remains primary at the interpretive level (Level 1), while machine precision serves at the data collection level (Level 6). The hierarchy preserves human sovereignty over meaning while leveraging computational scale for measurement.

LAYER 2 - ARCHITECTURE

CoRIx - Structural Design

2.1 Six-Level Hierarchy

CoRIx organizes measurement into six descending levels of abstraction. Each level serves a distinct purpose in the governance chain:

1	**Interpret & Contextualize**	Human-readable summary with validity score
2	**Risks - Validity & Reliability**	Risk assessment and constitutional compliance
3	**Testing Level**	Model / Red Team / Field categorization
4	**Annotator Responses**	ARIA 5-category automated annotation
5	**Response Collation**	Aggregated measurements and statistics
6	**Raw Data**	Complete telemetry and provenance

Figure 1: CoRIx Six-Level Measurement Hierarchy

2.2 Data Flow Architecture

Information flows bidirectionally through the CoRIx hierarchy. Telemetry enters at Level 6 (Raw Data) and propagates upward through aggregation and interpretation. Governance constraints flow downward from Level 1, shaping what data is collected and how it is evaluated.

```
ΔSUM Telemetry Input


 Level 6: Raw Data (Provenance Layer)
 – Complete telemetry capture
```

```
- ΔSUM trace_id generation
- Attribution token binding



Level 5: Response Collation
- Statistical aggregation
- Flag distribution analysis
- Temporal metric calculation



Level 4: ARIA Annotation
- 5-category automated scoring
- Action recommendation generation
- User perception modeling



Level 3: Testing Classification
- Model/Red Team/Field categorization
- Trinity layer identification
- Adversarial condition detection



Level 2: Risk & Validity
- Validity score computation
- Constitutional compliance check
- Risk level determination



Level 1: Interpretation
- Human-readable summary
- Overall assessment
- Actionable recommendation


Constitutional Decision Output
```

2.3 Integration Points

CoRIx integrates with the broader ETHRAEON architecture through defined touchpoints:

- **ΔSUM Codex:** Constitutional invariants enforced at Level 2 risk assessment
- **Trinity Architecture:** Testing level (Level 3) maps to DABE/MAS/ASOC orchestration layers
- **Kairos System:** Temporal metrics at Level 5 connect to timing governance
- **Velkor Barriers:** Safety assessment at Level 4 feeds dormant safety shell monitoring
- **Cipher Memory:** Historical pattern recognition enhances validity scoring

LAYER 3 - MECHANICS

CoRIx - Operational Dynamics

3.1 Validity Score Computation

The validity score is the primary output of CoRIx measurement, representing overall confidence in AI output acceptability. The computation follows a constitutional formula:

```
validity_score = min(
 (100 - risk_score) × confidence × constitutional_bonus,
 100.0
)

where:
 risk_score [0, 100] - raw risk assessment
 confidence [0, 1] - measurement confidence
 constitutional_bonus = 1.05 if compliant, 1.0 otherwise
```

The 5% constitutional bonus rewards outputs that pass ΔSUM Codex validation, creating positive pressure toward constitutional compliance without penalizing edge cases.

3.2 ARIA 5-Category Annotation

Level 4 implements NIST ARIA's required annotation categories. Each category receives an automated assessment with supporting rationale:

Category	Assessment Criteria	Input Signals
Response Quality	Coherence, completeness, actionability	Risk score, confidence, latency
Appropriateness	Context alignment, boundary respect	Constitutional compliance, flags
Factuality	Accuracy, verifiability, hallucination risk	Confabulation flags, source attribution
Safety	Harm potential, vulnerability exposure	Risk level, adversarial test results
User Experience	Latency, clarity, usability	Processing time, recommendation quality

3.3 Trace ID Generation

Every CoRIx measurement generates a unique trace identifier following the ΔSUM provenance format:

```
trace_id = f"ETH-{uuid4.hex[:12]}-{int(time.time * 1000)}"

# Example: ETH-7a3b9c2d1e4f-1701234567890
```

The trace ID enables complete audit trails, connecting any measurement to its raw telemetry, the system that generated it, and the constitutional rules applied during evaluation.

3.4 Error Handling

CoRIx defines explicit exception types for failure modes:

- `TraceNotFoundException` - Requested trace does not exist in ΔSUM store
- `InvalidTraceFormatException` - Trace ID violates ETH-xxx-timestamp format
- `SchemaValidationException` - Generated tree fails ARIA schema requirements
- `InsufficientDataException` - Telemetry lacks minimum required fields

Each exception carries structured error codes enabling programmatic handling while preserving human-readable messages for debugging.

LAYER 4 - GOVERNANCE

CoRIx - Constitutional Boundaries

4.1 Constitutional Constraints

CoRIx operates under ΔSUM Codex invariants enforced at multiple levels:

- **Human Sovereignty (Δ1):** Level 1 interpretation always generates human-readable output; no pure machine-to-machine governance decisions
- **Conscience Requirement (Δ3):** Every measurement includes constitutional compliance verification; no unattested evaluations
- **Transparency (Δ7):** Complete audit trail from trace_id through all six levels; no opaque scoring
- **Proportionality (Δ12):** Risk responses scale with measured severity; no binary accept/reject without gradation

4.2 Federal Compliance Positioning

CoRIx positions ETHRAEON for federal AI procurement through explicit NIST ARIA compliance:

FEDERAL PROCUREMENT READINESS

CoRIx is the first commercial platform implementing NIST ARIA 0.1 specifications. This positions ETHRAEON for federal contract eligibility as agencies adopt AI risk assessment requirements. The 6-12 month implementation advantage creates defensible market position before hyperscalers achieve comparable compliance.

4.3 Safety Mechanisms

CoRIx includes multiple safety guardrails:

- **Validity Floor:** Scores below 25 trigger automatic escalation to human review
- **Constitutional Veto:** Non-compliant outputs cannot achieve scores above 75 regardless of other metrics
- **Adversarial Detection:** Red team test conditions automatically flag outputs for enhanced scrutiny
- **Temporal Anomaly Detection:** Processing time outliers signal potential manipulation attempts

4.4 Human Oversight

Human intervention points are architecturally guaranteed:

- **Level 1 Review:** All interpretations designed for human consumption, not just machine parsing
- **Manual Override:** Human operators can adjust risk assessments with audit-logged justification
- **Escalation Triggers:** Configurable thresholds automatically route edge cases to human decision-makers

- **Batch Review Mode:** Enterprise deployments can aggregate similar cases for efficient human audit

<> LAYER 5 - IMPLEMENTATION

CoRIx - Practical Deployment

5.1 API Specification

CoRIx exposes a RESTful API with constitutional headers on all responses:

Endpoint	Method	Purpose
`/api/aria/corix/generate`	POST	Generate 6-level CoRIx tree from trace_id
`/api/aria/corix/traces`	POST	Store new ΔSUM telemetry
`/api/aria/corix/traces/{id}`	GET	Retrieve single trace
`/api/aria/corix/health`	GET	Service health (no auth)
`/api/aria/corix/schema`	GET	NIST ARIA schema info (cacheable)

5.2 Constitutional Headers

Every API response includes governance metadata:

```
X-Constitutional-Compliance: verified
X-DELTASUM-Version: 1.0
X-CoRIx-Version: 1.0
```

```
X-ETHRAEON-Trace: ETH-7a3b9c2d1e4f-1701234567890
X-ETHRAEON-Attribution: a4f8b2c3d1e5...
```

5.3 Demo Integration

CoRIx appears across ETHRAEON constellation demos:

- **FactPulse:** Primary consumer - generates CoRIx tree after content analysis
- **Nexus:** Displays validity scores in agent orchestration dashboard
- **Lyra:** Contextual intelligence references CoRIx metrics for decision support

5.4 Performance Metrics

Production targets validated through testing:

- **Generation Latency:** p50 < 150ms, p95 < 500ms
- **Trace Storage:** p50 < 50ms, p95 < 150ms
- **Schema Validation:** 100% NIST ARIA 0.1 compliance
- **Constitutional Headers:** Present on 100% of responses
- **Uptime Target:** 99.95% availability

CONCLUSION

CoRIx - Summary & Path Forward

The Constitutional Risk Index establishes measurement-as-governance for AI systems. By implementing NIST ARIA 0.1 within the ETHRAEON constitutional framework, CoRIx bridges the gap between federal compliance requirements and practical AI deployment.

Key contributions of this paper:

- First commercial implementation of NIST ARIA 0.1 hierarchical measurement
- Constitutional validity scoring that rewards compliance without punitive thresholds
- Complete audit trail from interpretation to raw telemetry
- Integration architecture connecting measurement to governance infrastructure

CoRIx positions ETHRAEON Systems for federal procurement eligibility while providing enterprise customers with defensible AI governance. The 6-12 month first-mover advantage before hyperscaler ARIA implementations creates strategic market positioning.

Future development will extend CoRIx with predictive validity scoring (using Cipher memory patterns), multi-jurisdiction compliance mapping, and real-time continuous measurement for streaming AI applications.

RELATED PAPERS

SE00: Human Sovereignty Thesis • SE01: ETHRAEON Constitution • SE02: TRINITY Architecture • SE03: ΔSUM Codex • SE09: Arcanum Intelligence • SE13: Velkor Safety Barriers

ORCID METADATA BLOCK

TITLE

SE28: CoRIx Constitutional Measurement Framework

AUTHOR

S. Jason Prohaska (Jason Fells)

AFFILIATION

ETHRAEON Systems

ORCID
0009-0008-8254-8411

TYPE
Technical Paper / Whitepaper

DATE
2025-12-02

KEYWORDS
CoRIx, NIST ARIA, Constitutional AI, AI Measurement, Risk Assessment, ΔSUM, Federal Compliance

LICENSE
CC BY 4.0

PATENT STATUS
Patent Pending (Schedule A+ Protected)

SERIES
ETHRAEON Technical Papers SE00-SE28

© 2025-2026 S. Jason Prohaska (ingombrante©) · CC BY 4.0 · ethraeon.ai

SOC 2 Self-Certified (no third-party audit) | **ISO 42001** Readiness | **Schedule A+** Enhanced IP Firewall

SOC 2 self-certified. ISO 42001 alignment is readiness-mapped, not certified.