TECHNICAL METHODOLOGY

AriaOS Validation Methodology

TRL 6 Verification Process for Governed Autonomous Intelligence

VERSION 2.5.3 **TRL 6 VALIDATED**

DOCUMENT TYPE	VERSION	TRL STATUS	DATE
Technical Methodology	2.5.3 Stable	TRL 6 Validated	February 2026

VALIDATION SCOPE

AriaOS has been validated across Apple Silicon (M1/M2/M3), x86_64 enterprise hardware (HP ProLiant G8), and Ubuntu 24.04 server environments with sustained stress testing, network degradation, and failure injection. Validation encompasses 800+ endpoints with 99.97% uptime under load and sub-50ms P95 latency.

ResilientMind AI LLC | SDVOSB | CAGE: 14JQ9 | UEI: NW3SNPP7QWF4 | SAM Registered joseph@resilientmindai.com

SECTION 1

Executive Summary

This document describes the validation methodology used to achieve Technology Readiness Level 6 (TRL 6) for AriaOS, a governed autonomous intelligence platform designed for denied, degraded, intermittent, and limited (DDIL) environments.

TRL 6 validation demonstrates that AriaOS operates effectively in relevant environments under realistic operational conditions. This methodology encompasses stress testing, chaos engineering, autonomous recovery verification, governance validation, and multi-platform compatibility testing across 800+ endpoints.

97.3%	99.97%	47ms	800+
Autonomous Recovery Rate	Uptime Under Load	P95 Recovery Latency	Endpoints Validated

Validation Scope: *AriaOS has been validated across Apple Silicon (M1/M2/M3), x86_64 enterprise hardware (HP ProLiant G8), and Ubuntu 24.04 server environments with sustained stress testing, network degradation, and failure injection.*

SECTION 2

TRL 6 Definition and Criteria

2.1 TRL 6 Requirements (NASA/DoD Standard)

Technology Readiness Level 6 requires demonstration of a system or subsystem model or prototype in a relevant environment. For AriaOS, this means:

- Representative operational environment (edge hardware, disconnected networks)
- Realistic mission scenarios (multi-agent coordination, recovery, governance)
- Sustained operational stress (hours to days of continuous execution)
- Failure injection and recovery validation
- Performance metrics under degraded conditions

2.2 AriaOS-Specific Success Criteria

Criterion	Target	Validation Method
Autonomous Recovery	>95% success rate	Chaos engineering with network/process failures
Governance Continuity	100% audit integrity	Sustained stress with policy enforcement verification

Criterion	Target	Validation Method
Multi-Agent Coordination	800+ endpoint mesh	Large-scale deployment with agent lifecycle testing
Recovery Time	P95 < 50ms	Latency measurement under degraded conditions
Memory Bus Stability	Zero corruption	Extended runtime with integrity verification

SECTION 3

Validation Methodology Phases

Phase 1: Platform Compatibility Validation

Objective: Verify AriaOS operates correctly on target hardware platforms.

Test Platforms:

- Apple Silicon: M1, M2, M3 processors (macOS 13+)
- x86_64 Enterprise: HP ProLiant G8 (Ubuntu 24.04 LTS)
- Edge Compute: Low-power ARM64 targets

Validation Steps:

1. Clean OS installation on target hardware
2. AriaOS deployment with default configuration
3. Functional verification of all core components (Context Kernel, Memory Bus, Agent Orchestration)
4. Performance baseline measurement (latency, memory usage, CPU utilization)
5. 24-hour stability run with monitoring

Success Criteria:

- Zero crashes or kernel panics
- All components operational
- Performance within acceptable bounds (P95 latency < 50ms)

Phase 2: Stress Testing and Load Validation

Objective: Validate system behavior under sustained operational stress.

Test Scenarios:

- API Endpoint Stress: 800+ concurrent endpoints with sustained request load
- Memory Bus Saturation: High-throughput message passing
- Agent Lifecycle Churn: Rapid agent creation, suspension, termination
- Governance Policy Enforcement: Complex policy trees under load

Load Profiles:

Load Type	Duration	Intensity
Baseline Load	6 hours	100 requests/sec
Peak Load	2 hours	500 requests/sec

Load Type	Duration	Intensity
Sustained Stress	48 hours	200 requests/sec
Spike Load	15 min bursts	1000 requests/sec

Metrics Collected:

- Request latency (P50, P95, P99)
- Memory usage and leak detection
- CPU utilization patterns
- Error rates and failure modes
- Recovery time after failures

Phase 3: Chaos Engineering and Failure Injection

Objective: Validate autonomous recovery and governance continuity under failure conditions.

Failure Injection Scenarios:

- **Network Degradation:** Packet loss (5%, 20%, 50%), latency injection (100ms, 500ms, 2s)
- **Process Termination:** Kill critical processes (agent orchestrator, memory bus, governance engine)
- **Resource Exhaustion:** CPU saturation, memory pressure, disk I/O limits
- **Dependency Failures:** Database unavailability, file system errors
- **Byzantine Failures:** Corrupted messages, invalid state transitions

Recovery Validation:

1. **Soft Recovery:** Agent self-healing, checkpoint restoration, state reconciliation
2. **Hard Recovery:** Component restart, supervisor escalation, system reboot
3. **Governance Verification:** Audit trail integrity, policy enforcement continuity
4. **Timing Analysis:** Recovery time measurement, cascading failure detection

Success Criteria:

- 95%+ autonomous recovery rate (no human intervention)
- Zero audit integrity loss (all events logged)
- Recovery time P95 < 5 seconds for soft recovery, < 30 seconds for hard recovery
- No cascading failures across independent components

Phase 4: Governance and Policy Validation

Objective: Verify deterministic governance enforcement and policy compliance.

Test Cases:

- **Weighted Voting:** Multi-agent consensus with configurable voting weights
- **Policy Trees:** Nested policy enforcement with AND/OR logic
- **Audit Integrity:** Chain-of-custody verification for all actions
- **Autonomy Level Enforcement:** L0 (HITL) through L4 (full autonomy) validation
- **Operator Override:** Human intervention and audit trail consistency

Validation Methods:

1. Policy compliance verification (100% enforcement rate expected)
2. Audit log inspection and integrity verification (hash chain validation)
3. Governance decision traceability (all decisions explainable)
4. Escalation pathway testing (autonomy level reversion under failure)

Phase 5: Multi-Platform Validation

Objective: Demonstrate platform independence and cross-hardware compatibility.

Platform	Configuration	Test Duration	Status
Apple Silicon (M1/M2/M3)	macOS 13+, 16GB+ RAM	200+ hours	Validated
HP ProLiant G8	Ubuntu 24.04, 32GB RAM	150+ hours	Validated
Edge Compute (ARM64)	Linux 6.x, 8GB+ RAM	Ongoing	In Progress

Cross-Platform Verification:

- Identical governance behavior across platforms
- Consistent recovery times (plus or minus 10% variance)
- Portable configuration and state
- Hardware abstraction layer validation

SECTION 4

Documented Validation Results

4.1 Published White Papers

- **Apple Silicon Stress Validation White Paper:** Comprehensive stress testing results on M1/M2/M3 platforms
- **HP G8 Chaos Validation:** Ubuntu 24.04 chaos engineering and failure injection results
- **SEI Technical Brief:** Software Engineering Institute technical review

4.2 Key Findings

Autonomous Recovery:

Metric	Result
Autonomous recovery success rate	97.3% across all failure scenarios
P95 soft recovery time	47ms
P95 hard recovery time	18 seconds
Data loss during recovery	Zero (100% audit integrity)

Performance Under Stress:

Metric	Result
Concurrent endpoints sustained	800+ for 48+ hours
P99 request latency under peak load	120ms
Memory stability	No leaks detected over 200+ hour runs
CPU utilization (normal / stress)	15-40% / 60-85%

Governance Validation:

Metric	Result
Policy enforcement compliance	100% (zero policy violations)
Audit trail integrity	100% (hash chain verification passed)
Weighted voting consensus	100% deterministic outcomes
Autonomy level enforcement	100% correct level transitions

SECTION 5

Test Environment Specifications

5.1 Hardware Configurations

Primary Test Bed (Apple Silicon):

Component	Specification
Processor	Apple M2/M3 (8-12 cores)
Memory	16-32 GB unified memory
Storage	512GB+ SSD
Network	Gigabit Ethernet + Wi-Fi 6

Enterprise Test Bed (HP ProLiant G8):

Component	Specification
Processor	Intel Xeon E5-2600 series (16-32 cores)
Memory	32-64 GB ECC RAM
Storage	RAID 1 SSD array
Network	Dual gigabit NICs

5.2 Network Conditions Tested

Condition	Parameters	Duration
Normal Connectivity	0% loss, <5ms latency	Baseline (hours)
Degraded Network	5-20% loss, 100-500ms latency	2-4 hours
Severely Degraded	30-50% loss, 500ms-2s latency	1-2 hours
Complete Isolation	100% external connectivity loss	4-8 hours
Intermittent Connectivity	Random 1-5 min outages	6-12 hours

SECTION 6

Metrics and Measurement

6.1 Performance Metrics

- **Latency:** P50, P95, P99 request/response times
- **Throughput:** Requests per second, messages per second
- **Resource Usage:** CPU, memory, disk I/O, network bandwidth
- **Stability:** Uptime, crash rate, memory leak detection

6.2 Reliability Metrics

- **Recovery Rate:** Percentage of failures autonomously recovered
- **Mean Time to Recovery (MTTR):** Average recovery time after failure
- **Audit Integrity:** Percentage of events successfully logged
- **Policy Compliance:** Percentage of policy-compliant actions

6.3 Governance Metrics

- **Decision Traceability:** Percentage of decisions with full audit trail
- **Voting Consensus:** Time to consensus, voting accuracy
- **Escalation Correctness:** Autonomy level transitions under failure
- **Human Override Response:** System response time to operator intervention

SECTION 7

Continuous Validation and Monitoring

7.1 Ongoing Validation

TRL 6 validation is not a one-time event. AriaOS undergoes continuous validation as new capabilities are added and as the system evolves.

Regression Testing:

- Automated stress test suite runs on every release candidate
- Chaos engineering scenarios executed weekly
- Performance benchmarks tracked over time
- Governance compliance verified with each policy update

7.2 Field Validation (TRL 7+ Path)

The next validation phase (TRL 7) requires demonstration in an operational environment. This will involve:

- Deployment in representative mission scenarios
- Integration with actual operational systems
- Extended field testing under real-world conditions
- Operator feedback and usability validation

SECTION 8

Validation Artifacts

8.1 Test Reports and Data

- Stress test logs and performance metrics
- Chaos engineering failure/recovery traces
- Governance audit logs and policy enforcement records
- Platform compatibility matrices

8.2 Published Documentation

- Apple Silicon Stress Validation White Paper (public)
- HP G8 Chaos Validation White Paper (public)
- SEI Technical Brief (public)
- Hardware Validation Targets (public)
- Degraded Modes Matrices (public)

8.3 Research Collaboration Artifacts

Research collaborators receive additional validation artifacts including:

- Raw test logs and traces
- Recovery timelines and failure analysis
- Governance records and audit chains
- Performance benchmark data

SECTION 9

Validation Standards and Compliance

9.1 Reference Standards

- **NASA TRL Scale:** Technology Readiness Levels 1-9
- **DoD TRA Handbook:** Technology Readiness Assessment guidance
- **ISO/IEC 25010:** Systems and software quality requirements
- **NIST AI RMF:** AI Risk Management Framework (reference)

9.2 Test Methodology Standards

- Chaos Engineering Principles (Netflix, Google SRE)
- Statistical significance testing (95% confidence intervals)
- Reproducible test environments (infrastructure as code)
- Independent verification (external research collaborators)

SECTION 10

Limitations and Future Work

10.1 Current Limitations

- **Scale:** Validated up to 800 endpoints; larger deployments require additional testing
- **Duration:** Longest continuous run: 200+ hours; extended multi-week runs pending
- **Hardware Diversity:** Additional ARM64 and NVIDIA platforms in progress
- **Field Conditions:** TRL 7 operational validation pending

10.2 Ongoing Validation Areas

- Extended duration testing (weeks to months of continuous operation)
- Large-scale mesh validation (1000+ endpoints)
- NVIDIA CUDA acceleration validation
- Heterogeneous platform mesh (mixed Apple/x86/ARM)

- Real-world mission scenario integration

AriaOS Validation Methodology
ResilientMind AI LLC | SDVOSB | CAGE: 14JQ9 | UEI: NW3SNPP7QWF4 | SAM Registered
Version 2.5.3 | February 2026

For validation artifacts or research collaboration inquiries: **joseph@resilientmindai.com**