NIST Cybersecurity Framework Profile for Artificial Intelligence

Public Comment Submission

Document: NIST IR 8596 (Preliminary Draft) **Submission Date:** January 17, 2026 **Comment Deadline:** January 30, 2026 **Submit To:** cyberaiprofile@nist.gov

---

Submitter Information

Organization: Vorion Risk, LLC **Contact:** Ryan Cason and Alex Blanc **Email:** contact@vorion.org **Website:** https://vorion.org **Relevant Standard:** BASIS (Behavioral Agent Safety and Interoperability Standard) **Standard URL:** https://basis.vorion.org

---

Executive Summary

Vorion Risk, LLC respectfully submits comments on NIST IR 8596, the Cybersecurity Framework Profile for Artificial Intelligence. We commend NIST's proactive approach to integrating AI-specific risks into the established CSF 2.0 framework.

Our comments focus on a critical gap in current AI governance frameworks: **operational efficiency and sustainability metrics for autonomous AI agents**. We propose that the Cyber AI Profile incorporate quantifiable efficiency requirements that complement its security-focused guidance, enabling organizations to govern AI systems holistically.

We offer the BASIS (Behavioral Agent Safety and Interoperability Standard) specification as a complementary technical framework that provides:

1. **Quantified trust scoring** for AI agent capability gating
2. **Resource manifests** aligned with cloud-native operational patterns
3. **Cost-to-value governance** with automatic throttling mechanisms
4. **Sustainability metrics** based on ISO/IEC 21031:2024 (SCI specification)

BASIS is an open standard (Apache 2.0 / CC BY 4.0) designed to integrate with NIST AI RMF and the Cyber AI Profile, providing the operational "efficiency layer" that current frameworks leave unspecified.

Detailed Comments

Comment 1: Incorporating Operational Efficiency into "Secure AI Systems"

Reference: Cyber AI Profile, AI Focus Area: "Secure AI Systems"

Current State: The profile appropriately emphasizes securing AI systems against adversarial threats, data poisoning, and model manipulation. However, it does not address the operational efficiency of AI systems as a security-adjacent concern.

Gap Identified: Inefficient AI agents represent a security risk through:

- **Resource exhaustion attacks** — Agents consuming excessive compute can create denial-of-service conditions

- **Cost amplification vulnerabilities** — Adversaries can trigger expensive reasoning modes for low-value tasks
- **Sustainability exposure** — Organizations face regulatory and reputational risk from uncontrolled AI energy consumption

Proposed Addition: We recommend the profile include guidance on operational efficiency controls as part of securing AI systems:

> *"Organizations should implement resource governance for AI systems, including:*
>
> - *Declared resource manifests specifying compute requirements and limits*
> - *Cost-to-value monitoring that tracks operational efficiency*
> - *Automatic throttling or degradation when resource consumption exceeds justified thresholds*
> - *Sustainability metrics aligned with ISO/IEC 21031:2024 (Software Carbon Intensity)"*

Supporting Framework: The BASIS Efficiency Specification (https://basis.vorion.org) provides a complete technical implementation including:

- Resource Manifest schema following Kubernetes dual-threshold patterns (requests/limits)
- Cost-to-Value (CTV) ratio algorithms with defined thresholds
- Degradation cascade: Alert (CTV > 2) → Throttle (CTV > 5) → Degrade (CTV > 10) → Stop (CTV > 20)
- Integration with the Green Software Foundation's Carbon Aware SDK

Comment 2: Quantified Trust Scoring for AI Agent Governance

Reference: Cyber AI Profile, mapping to CSF Govern and Manage functions

Current State: The AI RMF and Cyber AI Profile establish governance structures and risk management processes but do not prescribe specific mechanisms for quantifying AI system trustworthiness at runtime.

Gap Identified: Without quantified trust metrics, organizations cannot implement graduated capability controls. Binary allow/deny decisions are insufficient for agentic AI systems that require nuanced permission management based on demonstrated behavior.

Proposed Addition: We recommend the profile reference quantified trust scoring as a best practice for AI agent governance:

> *"For autonomous AI agents, organizations should consider implementing graduated trust mechanisms that:*
>
> - *Assign numeric trust scores based on historical behavior and compliance*
> - *Map trust levels to capability tiers with progressive permission unlocks*
> - *Apply trust decay for inactive agents to prevent stale high-trust entities*
> - *Amplify negative impacts from failures to incentivize reliable behavior"*

Supporting Framework: The BASIS Trust Model (https://basis.vorion.org/spec/trust-scoring) provides:

Tier	Score Range	Default Capabilities
Sandbox	0-99	Isolated testing only
Provisional	100-299	Read public data, internal messaging
Standard	300-499	Limited external communication
Trusted	500-699	External API calls

Certified	700-899	Financial transactions
Autonomous	900-1000	Full autonomy within policy bounds

The trust score algorithm includes:

- 7-day half-life decay for inactive agents
- 3x failure amplification multiplier
- Tier boundaries that prevent capability skipping

Comment 3: Addressing Agentic AI in "Conduct AI-Enabled Cyber Defense"

Reference: Cyber AI Profile, AI Focus Area: "Conduct AI-Enabled Cyber Defense (Defend)"

Current State: The profile addresses using AI for cyber defense but does not specifically address governance of autonomous AI agents that may take defensive actions without human approval.

Gap Identified: AI agents deployed for cyber defense introduce unique risks:

- **Autonomous action scope** — Defensive agents may block legitimate traffic or isolate systems
- **Escalation requirements** — Some defensive actions require human oversight
- **Audit requirements** — Defensive actions must be logged with cryptographic integrity for forensic analysis

Proposed Addition: We recommend explicit guidance for agentic AI in cyber defense contexts:

> "*When deploying AI agents for autonomous cyber defense operations, organizations should:*
>
> - *Implement capability gating that requires human escalation for high-impact defensive actions*
> - *Maintain immutable audit trails with cryptographic hash chains for all agent decisions*
> - *Define clear boundaries between autonomous and human-supervised defensive capabilities*
> - *Establish governance checkpoints before agents execute defensive actions*"

Supporting Framework: The BASIS four-layer architecture directly addresses this:

```
LAYER 1: INTENT   → Parse defensive action request, classify risk level
LAYER 2: ENFORCE  → Evaluate against trust score, apply escalation rules
LAYER 3: PROOF    → Create SHA-256 chained audit record
LAYER 4: CHAIN    → Optional blockchain anchoring for independent verification
```

Governance decisions include: ALLOW, DENY, ESCALATE (require human approval), or DEGRADE (reduced scope).

Comment 4: Efficiency Metrics for AI Sustainability Reporting

Reference: Alignment with EU AI Act environmental provisions and CSRD requirements

Current State: The Cyber AI Profile does not address AI system sustainability or energy efficiency, despite increasing regulatory pressure (EU AI Act Article 40, CSRD disclosure requirements) and enterprise sustainability commitments.

Gap Identified: Organizations deploying AI systems face growing requirements to report on:

- Energy consumption of AI workloads
- Carbon emissions associated with AI operations
- Computational efficiency relative to value delivered

Microsoft's 30% carbon footprint increase (2020-2023), largely attributed to AI infrastructure, demonstrates the materiality of this concern.

Proposed Addition: We recommend the profile include sustainability considerations:

> "*Organizations should establish metrics for AI system sustainability, including:*
>
> - *Energy consumption per functional unit (e.g., Wh per 1,000 queries)*
> - *Carbon intensity aligned with ISO/IEC 21031:2024 (SCI specification)*
> - *Embodied carbon attribution for AI hardware*
> - *Carbon-aware scheduling that shifts workloads to low-carbon periods*"

Supporting Framework: The BASIS Efficiency Specification incorporates the SCI formula:

```
SCI = ((E × I) + M) / R

Where:
  E = Energy consumption (kWh)
  I = Marginal carbon intensity (gCO2eq/kWh)
  M = Embodied carbon (gCO2eq)
  R = Functional unit (per API call, per session, etc.)
```

Critically, BASIS explicitly excludes market-based measures (carbon offsets, RECs) from efficiency scoring, focusing on actual emissions reduction.

Comment 5: Reasoning Mode Governance for Cost Control

Reference: Risk management for generative AI systems

Current State: The Generative AI Profile (NIST AI 600-1) addresses unique GenAI risks but does not specifically address the operational cost implications of reasoning-enabled models.

Gap Identified: Recent analysis (AI Energy Score v2, January 2026) found that reasoning-enabled models consume **150-700x more energy** than standard inference, with an average of 30x across tested models. Without governance:

- Agents may use expensive reasoning for trivial tasks
- Organizations lack visibility into reasoning mode cost
- No automatic controls exist to optimize reasoning usage

Proposed Addition: We recommend explicit guidance for reasoning mode governance:

> "*For AI systems with reasoning capabilities (chain-of-thought, extended thinking, etc.), organizations should:*
>
> - *Track reasoning mode usage separately from standard inference*
> - *Implement task classification to determine when reasoning is justified*
> - *Establish reasoning token budgets that constrain expensive operations*
> - *Enable automatic reasoning mode disabling for simple, routine tasks*"

Supporting Framework: BASIS defines reasoning mode governance including:

- Task classification (REASONING_JUSTIFIED vs REASONING_WASTEFUL task types)
- Reasoning budget tracking per agent
- Automatic model selection based on task complexity
- Separate CTV calculation for reasoning operations

Proposed Integration Path

We propose the following integration between BASIS and the NIST Cyber AI Profile:

NIST CSF Function	Cyber AI Profile Focus	BASIS Contribution
GOVERN	Establish AI governance	Trust tiers, capability taxonomy, policy constraints
IDENTIFY	Understand AI context	Resource manifests, hardware tier classification
PROTECT	Implement safeguards	Capability gating, escalation requirements
DETECT	Monitor AI systems	CTV monitoring, efficiency alerts, anomaly detection
RESPOND	Address AI incidents	Automatic throttling, degradation cascade
RECOVER	Restore AI operations	Recovery from auto-stop, trust score rebuilding

---

Technical Resources

The following open resources are available for NIST review:

Resource	URL	License
BASIS Specification	https://basis.vorion.org	CC BY 4.0
Core Specification	https://basis.vorion.org/spec/overview	CC BY 4.0
Efficiency Specification	https://github.com/voriongit/vorion/blob/master/basis-core/specs/BASIS-EFFICIENCY.md	Apache 2.0
Capability Taxonomy	https://basis.vorion.org/spec/capabilities	CC BY 4.0
JSON Schemas	https://github.com/voriongit/vorion/blob/master/basis-core/specs/BASIS-JSON-SCHEMAS.md	Apache 2.0
Compliance Mapping	https://github.com/voriongit/vorion/blob/master/basis-core/specs/BASIS-COMPLIANCE-MAPPING.md	CC BY 4.0
Reference Implementation	https://cognigate.dev	Apache 2.0

Conclusion

The NIST Cybersecurity Framework Profile for Artificial Intelligence represents an important step toward comprehensive AI governance. We believe the profile would be strengthened by incorporating:

1. **Operational efficiency requirements** that address resource consumption as a security-adjacent concern
2. **Quantified trust scoring** that enables graduated capability controls for autonomous agents
3. **Sustainability metrics** aligned with ISO/IEC 21031:2024 and emerging regulatory requirements

4. **Reasoning mode governance** that addresses the 150-700x energy cost differential

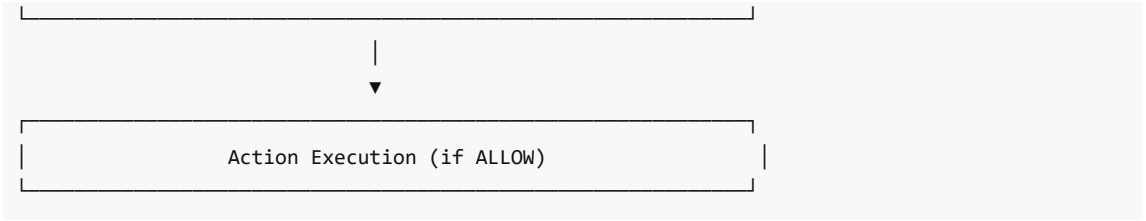
The BASIS standard provides ready-to-adopt technical specifications for each of these areas, designed to complement rather than compete with NIST frameworks. We welcome the opportunity to discuss integration approaches with NIST staff and contribute to the ongoing development of AI governance standards.

We are committed to supporting NIST's mission of promoting U.S. innovation and industrial competitiveness through responsible AI governance that balances security, efficiency, and sustainability.

Respectfully submitted,

Vorion Risk, LLC https://vorion.org https://basis.vorion.org

Appendix A: BASIS Architecture Overview



Appendix B: Cost-to-Value Governance Thresholds

CTV Ratio	Status	Automated Response
< 1.0	Excellent	Continue (value exceeds cost)
1.0 - 2.0	Acceptable	Monitor
2.0 - 5.0	Marginal	Alert agent, suggest optimization
5.0 - 10.0	Poor	Throttle (reduce to 50% capacity)
> 10.0	Unacceptable	Stop, require human review

Auto-Stop Conditions:

- CTV ratio > 10.0 sustained over rolling window
- 5+ consecutive operation failures
- Rolling average value score < 0 (negative value production)
- Resource consumption > 150% of declared manifest limits
- Carbon budget exhausted

Appendix C: Alignment with Existing Standards

Standard	BASIS Alignment
NIST AI RMF	BASIS provides quantifiable metrics for Measure function; efficiency controls for Manage function
ISO/IEC 42001	BASIS complements AIMS with operational performance certification
EU AI Act	BASIS addresses Annex IV documentation requirements; positions for August 2028 efficiency standards
ISO/IEC 21031:2024 (SCI)	BASIS adopts SCI formula as canonical sustainability metric
MLPerf Inference	BASIS efficiency tiers based on MLPerf methodology (90th percentile latency)
Kubernetes	Resource Manifests follow K8s dual-threshold pattern (requests/limits)

Document Version: 1.0 Prepared: January 17, 2026 For submission to: cyberaiprofile@nist.gov by January 30, 2026