How the Research Was Done

Methodology for the AI Cyber Security Research Boardroom Study

AI Cyber Security Research Boardroom

2026-05-31

Contents

1 How the Research Was Done AICSR-METHOD-2026-001

Field	Value
Reference key	`method-ref`
Availability	AVAILABLE
Capture method	`primary`
Source type	local_repository
URL / path	output/How-The-Research-Was-Done.md
Captured (UTC)	2026-06-23T03:31:52Z
Content type	text/markdown
HTTP status	n/a

1.1 Boardroom Citation Context

Reader-facing methodology; simulation disclosure.

1.2 Source Location

output/How-The-Research-Was-Done.md

1.3 Captured Content

2 How the Research Was Done

Document ID: AICSR-METHOD-2026-001
Version: 1.1
Publication Date: 2026-05-31

This document describes **how** the Cyber-Security and AI Diligence research was conducted. It is written for readers who want to understand the approach behind the study reports, mitigation strategy, session transcripts, and reference materials — without access to internal tooling or generation instructions.

2.1 1. Purpose

The research set out to answer a practical question: *What does a defensible minimum program look like for organizations deploying AI systems in environments where cybersecurity risk, supply-chain exposure, and government-adjacent infrastructure pressures are all material?*

The work product is a structured set of findings, frameworks, and recommendations — chiefly the **Minimum Viable AI Diligence Program (MVAP)** — produced through sustained expert deliberation rather than a single-author literature review.

2.2 2. Research Topic and Guiding Questions

2.2.1 Primary topic

Cyber-Security and AI Diligence Research — the intersection of enterprise AI adoption, security control design, open-source zero-day risk, and U.S. government infrastructure governance.

2.2.2 Base questions

Deliberation was organized around one primary topic and multiple sub-questions that emerged as consensus formed. Initial guiding questions included:

1. What pillars and controls constitute a **minimum viable** AI diligence program for enterprise deployment?
2. How should organizations treat **open-source and AI gateway** components that sit on exploitation timelines measured in hours, not quarters?
3. To what extent can enterprises rely on **federal cybersecurity infrastructure** (CISA, KEV, classified-contractor oversight) when building their own AI risk programs?
4. How should **red-team and blue-team** evidence change control priority when theoretical risk meets operational exploit paths?
5. When should a specification increment (e.g., MVAP v1.1 v1.2) be adopted, and what dissent should remain on record?

Sub-questions were introduced sequentially as the panel resolved or deadlocked on prior items. The full sequence of questions and answers is preserved in the session transcript compendium.

2.3 3. Expert Panel

The research employed a **team of experts** — thirty-one specialist roles organized into complementary domains:

Domain	Representative expertise
Governance	Moderation and independent fact verification (non-voting)
Compliance	Regulatory obligation, liability, auditability
Senior security architecture	CISSP-tier strategy, identity, and system design
Security operations	SSCP-tier hands-on detection and response
Entry-level security	Fresh operational perspectives and implementation friction
Vulnerability research	Zero-day discovery and exploit feasibility
Offensive security	Practical exploitation and attack-chain construction
Red rapid response	Time-pressured offensive pressure-testing
Blue rapid response	Time-pressured defensive countermeasures

Twenty-seven participants held voting seats. Two governance roles (moderator and court reporter) facilitated and verified but did not vote.

Each expert was represented by an **independently configured profile** capturing title, domain expertise, career perspective, communication style, and known biases. Profiles are on file under `participants/` for inspection.

2.4 4. Deliberation Model

Research was conducted as a **multi-agent expert debate**, not a single monolithic analysis.

2.4.1 Separate expert agents

Each profile operated as a **separate agent** with its own subject-matter lens. Agents were not collapsed into one voice; each spoke in turn from its defined expertise.

2.4.2 Structured debate

Given the topic and active sub-question, each voting expert was asked to articulate:

- Arguments **in favor** of a position or control
- Arguments **against** it, or risks that would make adoption impractical
- A synthesized **position** and **recommendation**

This forced explicit tradeoff analysis rather than unanimous checklist approval.

2.4.3 Cross-examination and position change

Agents did not deliver isolated monologues. Each turn occurred **after others had spoken** in the same round. Experts were expected to:

- Weigh peer arguments against their own domain knowledge
- Acknowledge compelling evidence from other specialties
- **Revise or reaffirm** their position when cross-domain reasoning warranted it

Where an expert changed stance between rounds, the transcript records the shift and the rationale tied to their expertise (for example, a compliance officer accepting a technical control after blue-team evidence, or an architect dissenting after red-team demonstrated exploit feasibility).

2.4.4 Moderation and verification

A moderator framed questions, tracked consensus, and advanced the agenda when majority agreement was reached. A court reporter maintained the authoritative transcript and audited factual claims against external sources. Unverified or speculative statements were labeled accordingly in the verification ledger.

2.5 5. Sessions and Chronology

Deliberation unfolded across **eleven recorded sessions** within a compressed **May 2026 simulation window** (chronology consolidated for readability). Sessions included:

Simulated date	Session focus
2026-05-01	Expert introductions and research framing
2026-05-02	Twenty rounds of pillar and control debate
2026-05-04	Red-team / blue-team exercise (Pillar 2 and 4)
2026-05-05	Remediation validation (pass 1)
2026-05-08	Maturity level L2 promotion review
2026-05-09	Remediation validation (pass 2)
2026-05-12	MVAP v1.1 adoption vote
2026-05-15	Zero-day tabletop (P7-05)
2026-05-16	Quarterly government risk review (Q1)
2026-05-20	MVAP v1.2 adoption vote
2026-05-21	Quarterly government risk review (Q2)

Complete transcripts: `sessions/` and the compiled dialog report `output/Boardroom-Complete-Dialog-Transcript.md`.

2.6 6. Voting and Recorded Conclusions

Formal decisions required a **majority of voting participants (14 of 27)** unless otherwise noted. Every ballot recorded:

- The **exact question** voted on
- **Per-participant vote** (YES / NO / ABSTAIN) when captured in transcript
- **Aggregate tally** (e.g. 18/27, 20/27)
- **Dissent rationale for each dissenter** — tied to that expert's domain, not generic objection

2.6.1 Selected recorded votes

Decision	Result	Outcome
NIST AI RMF as MVAP governance floor	**18/27**	PASSED — Pillar 1 governance baseline
OWASP LLM Top 10 for customer-facing apps	**20/27**	PASSED — Pillar 2 application security
KMS encryption for weights/embeddings at rest	**18/27**	PASSED

Decision	Result	Outcome
Dual compliance model (regulatory floor + pipeline ceiling)	**18/27**	PASSED
MVAP L2 operational promotion	**18/27**	PASSED — maturity review
MVAP v1.1 adoption	**20/27**	PASSED — P6 firmware + P7 zero-day elevated
MVAP v1.2 — P7-10 KEV SLA	**19/27**	PASSED
MVAP v1.2 — P7-11 gateway hardening	**21/27**	PASSED
MVAP v1.2 — P6 tier-2 firmware	**18/27**	PASSED
MVAP v1.2 — P7-09 transitive SBOM	**17/27**	CONDITIONAL — lowest majority item
Government risk register reaffirmation (Q2)	**20/27**	PASSED — GOV-02 remains Critical+

2.6.2 Dissent record with rationale (MVAP v1.1 adoption)

Participant	Vote	Rationale
Marcus Thorne	**NO**	Documentation and audit burden on small teams outweighs near-term P7 benefit without automated evidence collection
NullByte	**NO**	SLSA Level 3 still absent from specification — supply-chain integrity gap unacceptable for tier-1 AI gateways
Kira Okonkwo	**NO**	24-hour KEV patch SLA not mandated; leaves operational teams exposed to sub-36-hour exploitation windows

Additional dissent on MVAP v1.2 items appears in adoption session transcripts (`sessions/2027-06-20-mvap-v1-2-adoption.md`). Formal ballot tables with **per-dissenter ra-**

tionale are maintained in `sessions/VOTE-RECORD.md`. Session conclusions link to this record; the dialog compendium preserves per-participant positions across rounds.

2.7 7. Evidence and External Sources (Highly Detailed)

Every source cited in deliberation, footnotes, or the verification ledger underwent **end-to-end evidence handling**: verification, download where possible, attached article files, and indexing.

2.7.1 Per-reference workflow

Step	What was done
Identify	Each citation assigned a unique footnote key (e.g. `nist-airmf`, `gao-classified`)
Verify	URL or repository path resolved; claim text compared to retrieved content
Download	Full text or substantial excerpt captured (up to 12,000 characters per article)
Attach	Three formats produced per reference: **Markdown, LaTeX, and PDF**
Fallback	When primary URL blocked (HTTP 403, WAF, JS-only page, 404), corroborating sources that reference the same fact were located and captured
Explain	Unavailable or partial captures document why retrieval failed and which mirror was used
Index	Entry registered in availability report and category materials index

2.7.2 Attached reference files

For each of **34** footnoted and verification-ledger sources:

```
output/references/articles/{key}.md
output/references/articles/{key}.tex
output/references/articles/{key}.pdf
```

Articles include: metadata table, boardroom citation context, captured content, and — when applicable — **Capture Provenance** (primary failure, fallback URL, recovery method).

2.7.3 Availability summary

Status	Meaning
Available	Substantial text captured from primary or mirror source
Partial	Source reached; extraction incomplete

Status	Meaning
Unavailable	Primary blocked; article explains why and cites corroborating mirror if found
Fallback recovered	Primary blocked; content captured from alternate source (press summary, GitHub data export, encyclopedia citing primary)

Index reports:

- `output/references/REFERENCE-AVAILABILITY-REPORT.md` (and `.tex`/`.pdf`) — per-source availability
- `output/RESEARCH-MATERIALS-INDEX.md` (and `.tex`/`.pdf`) — **all research materials by category**

2.7.4 Fallback examples (primary blocked)

Source	Primary blocker	Corroborating capture
GAO-26-107861	gao.gov HTTP 403	Bloomberg Government summary
GAO-26-109159 (water sector)	gao.gov HTTP 403	Route Fifty GAO summary
MITRE ATLAS website	JavaScript SPA	`mitre-atlas/atlas-data` YAML export
NJCCIC advisory	Imperva Incapsula WAF	Wikipedia Salt Typhoon (cites Financial Times)
Sigstore cosign docs	Legacy URL 404	Current `/cosign/` documentation path

2.7.5 Verification ledger

`sessions/verification-ledger.md` classifies every audited claim:

- **Verified** — footnote key links to archived article with supporting text
- **Partial** — evidence incomplete; article path noted
- **Unverified** — no supporting capture
- **Projected speculation** — analytically useful; not externally confirmed

---

2.8 8. Research Outputs

Output	Location	Description
Comprehensive study	`output/Cyber-Security-AI-Diligence-Research-Study.md`	AI Cyber Security Diligence Research Study (AICSR-STUDY-2026-001)

Output	Location	Description
Mitigation strategy	`output/MVAP-Complete-Mitigation-Strategy.md`	Crate-by-crate implementation roadmap
Dialog compendium	`output/Boardroom-Complete-Dialog-Transcript.md`	Full session transcript
Abstracts	`output/Boardroom-Comprehensive-Abstracts.md`	Per-document and per-session summaries
Reference archive	`output/references/`	Footnoted source captures (MD/LaTeX/PDF per key)
Materials index	`output/RESEARCH-MATERIALS-INDEX.md`	All artifacts listed by category
Methodology	`output/How-The-Research-Was-Done.md`	This document
Vote record	`sessions/VOTE-RECORD.md`	Formal ballots with dissent rationale
Project metadata	`PROJECT.md`	Name, topic, layout, document IDs
MVAP specifications	`mvap/`	Adopted and draft framework versions
Participant profiles	`participants/`	Expert role definitions
Session transcripts	`sessions/`	Primary deliberation record

2.9 9. Limitations and Disclosure

This research was conducted as a **structured multi-agent simulation**. No physical boardroom convened. Expert voices, debate exchanges, votes, and exercises were produced through configured specialist profiles deliberating under moderation and court-report verification.

Readers should treat:

- **Verified facts** (footnotes and ledger-approved claims) as externally anchored
- **Deliberative consensus** as structured expert judgment under time and scope constraints
- **Speculative projections** as labeled hypothesis, not forecast

All underlying materials remain in the repository for independent inspection. This methodology document intentionally omits internal generation instructions, prompts, and operational tooling detail.

2.10 10. Summary

The research combined a **defined topic and evolving base questions**, a **multi-disciplinary expert panel** represented as **separate agents**, and **iterative debate** in which participants responded to one another and **updated positions** when cross-domain evidence warranted. **Votes and dissent** were recorded in session conclusions and adoption ballots. Synthesis reports distill those deliberations into frameworks and recommendations; transcripts and profiles preserve the underlying record.

Archived reference article for AICSR-STUDY-2026-001 footnote corpus.