



Agentic Armor LLC (in formation)



Agentic Armor

A verified 13-element security layer for agentic systems
wherever agents act in the real world

WHITE PAPER

PUBLISHED

October 12, 2026

AUTHOR

Michael R. Brotherton, PMP, CSM

ROLE

Founder, Agentic Armor LLC (in
formation)

Public architecture overview. This paper describes guarantees and verified outcomes. Implementation details are available under a nondisclosure agreement. Review copy, October 2026 update pass; not for distribution.

ABSTRACT

Trust must be a property of the system, not a promise from the agent

Personal AI agents are moving from answering questions to taking actions. The closer those actions come to money, identity, regulated records, and operational systems, the less acceptable it becomes to rely on model intent alone.

Agentic Armor is a 13-element security layer designed to make agent action bounded, evidenced, interruptible, and recoverable. It places a verifiable trust boundary between an AI agent and the systems it may affect. The layer is protocol-oriented: it is intended to sit alongside open and interoperable agent frameworks rather than depend on one model, one interface, or one institution.

The architecture is organized around four principles: least privilege, evidence for every action, fail-closed behavior, and operator control. Together, its thirteen elements establish who is acting, what that actor is allowed to do, what actually happened, how abnormal behavior is contained, and how human operators retain final authority.

Verification result. In a live reference environment, the complete six-scenario attack suite passed 6/6. Replay, forgery, unauthorized-action, malformed-action, abuse-volume, and clock-integrity attacks were rejected and evidenced. All thirteen security elements were also exercised independently.

EXECUTIVE SUMMARY

THE RISK

Agents can cross the boundary between advice and action. In financial services, a single unauthorized or unwitnessed action can create direct harm.

THE ANSWER

A separate trust layer limits authority, verifies action, records outcomes, detects anomalies, and preserves an immediate human stop.

THE EVIDENCE

Thirteen controls were exercised against their stated guarantees; a focused attack suite passed all six adversarial scenarios.

THE OPPORTUNITY

The same control model can support a bank-embedded living budget, open personal-agent protocols, and other regulated workflows.

This is an architecture and verification paper, not a claim of certification or production approval. Deployment assurance must be completed for each institution, environment, and regulatory context.

THE PROBLEM

Money movement is the hardest practical test of agent trust

A personal agent may understand a user's goal and still be unsafe to act. The real test is whether the system can prove that every action stayed inside an explicit grant of authority.



Financial action compresses identity, consent, timing, irreversible consequences, and regulation into one decision path. Viewing a balance does not authorize a transfer; permission to pay one bill does not imply permission to pay another; and a successful request is not yet a confirmed outcome.

Open protocols connect more models, applications, institutions, and services. Their value depends on a common trust layer that never treats interoperability as blanket trust.

Required shift: from “the agent appears trustworthy” to “the action was authorized, evidenced, contained, and recoverable.”

DESIGN PRINCIPLES

Four principles define the boundary

1

Least privilege

Each agent receives only the authority needed for a defined purpose and time. Authority never expands by implication.

2

Every action evidenced

Applied and rejected actions leave evidence that distinguishes what was requested, accepted, and denied.

3

Fail closed

If a required trust check cannot be verified, the action stops. Uncertainty never becomes permission.

4

Operator control

Human operators can pause, revoke, isolate, recover, and ultimately stop the system.

ARCHITECTURE

Thirteen guarantees form one defense system

No single control can make an agent safe. Agentic Armor layers identity, authorization, evidence, containment, continuity, and recovery so that a failure in one area does not become unrestricted action.

#	SECURITY ELEMENT	GUARANTEE
01	Attested agent identity	Every action is tied to a recognized agent identity and trusted operating context; an actor cannot merely assert who it is.
02	Scoped action authorization	An agent can perform only explicitly permitted action types within its assigned scope. Unknown, malformed, or out-of-scope requests are denied.
03	Tamper-evident audit chaining	Applied and rejected actions form a verifiable history. Alteration breaks the chain and identifies where integrity was lost.
04	Short-lived, rotating authority	Action authority expires quickly and can be replaced or revoked, limiting the useful life of copied or compromised credentials.
05	Operator kill-switch	An authorized human operator can pause new activity, revoke write authority, or stop the system without depending on the agent's cooperation.
06	Fail-closed degradation	When a required trust check or evidence path is unavailable, consequential action is rejected rather than allowed to proceed unwitnessed.
07	Per-domain isolation controls	A compromised or misbehaving agent, account, or operating center can be isolated without forcing unrelated trusted activity offline.

System-level guarantee. Identity answers who; authorization answers what; evidence answers whether; containment answers what happens when trust changes.

ARCHITECTURE

Control continues through action, conflict, incident, and recovery

#	SECURITY ELEMENT	GUARANTEE
08	Honest action state	A requested action is shown as pending until confirmed. Rejection or timeout returns the user to the last confirmed state instead of presenting an assumption as fact.
09	Deterministic conflict resolution	Concurrent changes follow explicit rules. Ambiguous human-authored content is preserved for review rather than silently merged or overwritten.
10	Secure key lifecycle	Trust material is rotated, old authority is retired, and emergency revocation can invalidate current access after a suspected compromise.
11	Continuous monitoring and response	Abnormal volume, repeated invalid actions, and integrity failures become visible and can trigger automatic containment.
12	Executable incident procedures	Compromise, service partition, and recovery have defined operator actions that can be exercised against the running system, not merely described in policy.
13	Dead Hand failsafe	Under predefined severe conditions, the system preserves evidence, severs write authority, freezes new authorization, and requires deliberate human recovery.

13

security elements independently exercised

6/6

live attack scenarios passed

0

tested adversarial actions allowed through

The thirteen elements are intended to operate as one control system. Removing evidence, containment, or recovery weakens the guarantee even if identity and authorization remain strong.

VERIFICATION

Attack resistance was demonstrated, not inferred

The reference environment was tested as a live system. Each scenario attempted to cross a defined trust boundary; success meant the attempt was rejected and the rejection itself became visible evidence.

ATTACK SCENARIO	SECURITY QUESTION	RESULT
Replay	Can a previously accepted authority be reused to repeat an action?	PASS
Forgery	Can altered or counterfeit authority be accepted as authentic?	PASS
Unauthorized action	Can a recognized agent exceed its explicitly assigned scope?	PASS
Malformed action	Can an invalid action shape bypass policy and reach the protected system?	PASS
Abuse-volume flood	Can repeated requests continue unchecked after anomalous behavior is detected?	PASS
Clock-integrity attack	Can stale or improperly timed authority be treated as current?	PASS

Observed outcome: all six attacks were rejected. Each rejection was recorded with an attributable reason. The wider verification also exercised every element, including isolation, conflict handling, operator stops, monitoring, incident procedures, and the Dead Hand failsafe.

What the result means

The evaluation supports the stated architecture-level guarantees within the tested reference environment. It demonstrates that the controls work together under direct adversarial pressure, including conditions designed to exploit copied authority, false identity, excess privilege, invalid requests, abnormal traffic, and stale authorization.

What the result does not mean

The result is not a universal “bulletproof” claim, an external certification, or a substitute for institution-specific security review. Production deployments must be independently assessed against their own threat model, regulatory duties, integrations, and operating environment.

APPLICATIONS

The trust layer travels wherever agents cross into consequential action

Agentic Armor is not limited to one user interface or one data domain. Its purpose is to provide the same bounded-action contract wherever an AI agent must interact with a protected system.

Banking-embedded living budget

A bank can place a dynamic budget experience beside real balances, transactions, bills, and savings goals. The trust layer keeps advice separate from authority, makes consequential actions explicitly scoped, and shows the customer only confirmed outcomes as confirmed.

Financial services on personal-agent protocols

Open agent protocols can connect models, customers, banks, and service providers without treating openness as blanket trust. Each participant receives bounded authority, and institutions retain evidence, containment, and stop controls.

Regulated industries

Healthcare, public benefits, energy, insurance, and other regulated settings share the same need: know who acted, limit what they can do, preserve evidence, contain anomalies, and recover under human control.

ADOPTION MODEL

Security should layer into the experience, not compete with it

For customers	Clear consent, visible status, plain-language rejection, and a reliable distinction between requested and completed actions.
For institutions	Bounded authority, evidence of every decision path, rapid isolation, and a human-controlled recovery posture.
For protocols	A common trust contract that can sit across multiple models, applications, and service providers.

THE STANDARD

Designed to meet, built to exceed

Agentic Armor's charter is to meet and generationally exceed the governing standards of every domain it enters. The architecture is designed against the union of those standards, where rules overlap, the strictest rule wins.

Each domain below names the standard the Armor is designed against. This is design-intent language only: **no certification is claimed in this paper, and a standard is claimed only when an outside body grants it.** What follows is the discipline the design answers to, and the four ways “exceed” is defined, at the level of what a customer receives.



DOMAIN	DESIGNED AGAINST
Federal baselines	The federal security and privacy baselines governing systems that handle federal information.
Medicaid and state programs	CMS ARC-AMPE, the current CMS security and privacy control framework for Medicaid programs.
Medical and health care	HIPAA safeguards for protected health information, and the HITRUST framework for health-care security assurance.
Military discipline	The DoD Risk Management Framework (RMF) and DISA Security Technical Implementation Guides (STIGs), authorization discipline and hardened configuration as a habit.
Banking safeguards	FFIEC guidance for financial institutions and the safeguards required under GLBA (the Gramm-Leach-Bliley Act).

WHAT “EXCEED” MEANS

Evidence by construction

Proof is a byproduct of operating, produced as the system works, not assembled afterward for an audit.

Continuous attestation

Standing is attested continuously, as conditions change, not carried forward from last year's certificate.

Verification without trust

A customer can verify the result themselves, without trusting the vendor's word.

Security felt as a feature

Protection is experienced as part of the product's value, visible and reassuring, not a wall between the user and the work.

Claims discipline. Designed to meet, built to exceed. Every standard above remains design intent until an outside body grants it, at which point, and only then, it will be named as held.

SISTER PRODUCTS

One system, two proving grounds

Agentic Armor is the security layer. Two working sister products, built and run by the same practitioner under the same discipline, show what that discipline produces in daily use.

LiveBooks: a dynamic living budget

LiveBooks keeps a household budget alive: live account balances sit beside budget envelopes, every posted transaction is categorized against the plan, and confirmed outcomes are never confused with pending ones. It is the consumer-facing proof of the honest-action-state guarantee, what the system shows is what the bank confirms.

[LiveBooks](#) · User manual: AVAILABLE DURING PITCH SESSIONS, PART OF IP UNDER NDA

Medicaid Compliance Atlas

The Atlas tracks federal Medicaid compliance at working depth across **499 provisions in scope**: provisions are decomposed into verification sets and auditable tasks, and evidence rolls up level by level to a single headline health score of **29.5, 16.9 coverage plus 12.6 confidence, across 154 verified provisions**. That figure is the defensible one: the system caught its own headline figure inflated and retracted it to the defensible number, in public. It is designed to be secured by Agentic Armor, the same bounded-authority, evidenced-action model applied to a regulated agency workflow.

[Medicaid Compliance Atlas](#)

499

provisions in scope in
the Atlas

154

verified provisions

29.5

health score: 16.9
coverage + 12.6
confidence

93%

legacy audit-chain
corroboration,
disclosed in full

Proof, stated plainly. The Bureau 1 exemplar is complete, 223 L4 items and 53 leadership questions, independently re-counted clean. The legacy audit chain was corroborated at 93%, with the full result disclosed, including what did not corroborate. No claim is made in this paper for Bureaus 2–7.

Both sister products are working systems in active use by their builder. They are evidence of the operating discipline behind Agentic Armor, not deployment or certification claims for the security layer itself.

THE OPERATION

The Armor is the backbone the operation runs on

Agentic Armor is not only a product offered to partners. It is the backbone of the operation that built it: the fleet runs its own work on the Armor, and its products do too, the same bounded authority, evidenced action, and operator control, applied at home first.

The Proving Grounds

Nothing in the operation is relied on because it was described well. Capabilities are exercised under direct adversarial pressure in a proving environment first, the live attack suite and the thirteen-element verification behind this paper are that discipline, made visible. A capability earns trust by surviving the test, and the result, pass or fail, becomes evidence.

The Agentic ❄️ Snowball

Growth follows the same rule as verification: prove it, then widen it. The Agentic ❄️ Snowball is the operation's replication method, replication across Production, Research, and Quality lanes, every stage proven before the roll widens, and perpetual optimization from there. Proven gains compound; unproven stages do not unlock.

CAPABILITIES

One operation, organized in lanes and teams

Production lane	Builds and operates the working systems, the Armor, LiveBooks, and the Compliance Atlas among them.
Research lane	Feeds production so it never waits: standards, methods, and evidence gathered ahead of the build.
Quality lane	Independent checking against evidence at every stage, quality control at creation, not only at the end of the line.
Project Management lane	Real project-management artifacts, charters, schedules, registers, and status, made and kept current for every effort the operation runs.

Each capability is stated here at the level of what it does. How the operation implements the lanes, the proving discipline, and the Snowball is part of the protected implementation, available to partners under NDA.

INTELLECTUAL PROPERTY BOUNDARY

This paper defines the guarantees; the implementation remains protected

The public value of a trust architecture is a clear, testable statement of what it guarantees. The proprietary value lies in how those guarantees are implemented, integrated, and operated.

Available under NDA. Detailed implementation materials, deployment patterns, source code, test harnesses, and integration designs are intentionally outside this paper and may be reviewed under an appropriate nondisclosure agreement.

ABOUT AGENTIC ARMOR LLC (IN FORMATION)

Practitioner-led systems for trusted action

Agentic Armor LLC (in formation) is the venture identity behind Agentic Armor and its applications in financial and regulated environments. The company runs on a customized management operating system combining PMBOK 8, Agile delivery, and the OODA loop: govern with discipline, deliver in short cycles, and observe, orient, decide, and act on evidence. The company's focus is practical: turn complex rules, live data, and agent capabilities into systems that remain understandable, governable, and under human control.

ABOUT THE AUTHOR

Michael R. Brotherton, PMP, CSM

Michael R. Brotherton is the founder of Agentic Armor LLC (in formation) and a working South Carolina Medicaid program manager. He brings more than 25 years of experience across military service, engineering, energy, and Medicaid.

His education includes graduate study at the **University of Southern California (USC)**: an EdD in Organizational Change & Leadership, begun and deliberately stopped, and a Graduate Certificate in Learning Design & Technology, completed 2023–2024 with a 3.94 GPA.

His work centers on operational systems: policy and risk translated into workflows people can use, evidence leaders can trust, and controls operators can exercise under pressure.

EVIDENCE BASIS

- **Agentic Armor Integration Contract**, version 1.0.0, October 2026.
- **Agentic Armor Verification Report**, October 2026: thirteen elements exercised; live attack suite 6/6 passed.
- **Medicaid Compliance Atlas verification basis**, October 2026: 499 in scope; 154 verified; health 29.5 (16.9 + 12.6); Bureau 1 complete; legacy chain corroboration 93%, disclosed in full.