ARCHITECTURAL OVERVIEW AND VERIFICATION PROTOCOL · ENGLISH REFERENCE EDITION

Proven Action. Power and Control over AI Agents: Strategy 2027–2035

Aleksandr Fishchuk (ORCID 0009-0000-6549-6113) · Scientific Director, Research Institute of System Synthesis · Full Russian monograph: 856 pages (16 October 2026).

DECISION LATENCY	ATTACK BENCHMARK	RECEIPT CHAIN	FULL EDITION
1.08 ms	157 / 157	5,094	856 pp.
Median Z3 solver check on CPU before tool execution	114 attack vectors and 43 routine calls, zero misses	SHA-256 linked receipts anchored via OpenTimestamps	35-page memorandum and 60-page excerpt in open access

1. Central engineering problem and thesis

Autonomous artificial intelligence (AI) agents execute database writes, code deployments, regulatory filings, and financial transfers directly inside enterprise software. Under civil, corporate, and regulatory law, a software model has no legal personality; liability for any destructive command or data leak rests entirely with the operating organization and its officers.

Probabilistic text filters and secondary language-model guardrails evaluate surface phrasing rather than formal preconditions, remaining vulnerable to indirect prompt injection and multi-step framing. The monograph formulates Information Security 3.0 (IS 3.0): placing a deterministic pre-execution gate (MITR) between the agent and target tools. Every proposed tool call over the Model Context Protocol (MCP) or application programming interface (API) is translated into quantifier-free linear integer arithmetic with uninterpreted functions (QF_UFLIA) and checked by the Z3 satisfiability modulo theories (SMT) solver before execution.

2. Structure of the 856-page monograph

VOLUME SECTION	SUBJECT MATTER AND ARCHITECTURAL CONTRIBUTION	VERIFIABLE ARTIFACT
Part I · Horizon 2026–2035	Transition from conversational assistants to tool-calling enterprise agents; taxonomy of real-world database, repository, and pricing failures.	Dated forecast registry · isslab.ai/book/
Part II · Cybernetics	Application of Ross Ashby's Law of Requisite Variety, Stafford Beer's Viable System Model (VSM), and the theory of inventive problem solving (TRIZ) to agent governance.	Formal soundness separation
Part III · MITR Gate	Zero-trust credential custody, Z3 SMT policy contracts, unsatisfiable core refusal feedback, and append-only SHA-256 / GOST R 34.11-2012 receipt chains.	20 enforce contracts · 6,414 code tests
Part IV · INDRA Bench	Independent adversarial verification of web and API surfaces with pre-egress destructive probe interception across seven vulnerability classes.	14/14 detections · indra.center
Part V · Law & Capital	Allocation of director liability, European Union Artificial Intelligence Act compliance (Articles 9, 12, 14, 15), ISO/IEC 42001, and algorithmic underwriting.	2 patent filings (06.2026, 09.2026)

INSTRUMENTAL GROUNDING OF THE MONOGRAPH

3. Empirical measurements and zero-trust verification protocol

Every quantitative claim in *Proven Action* is bound to a dated server measurement and an open verification script. The guarantor of MITR is mathematical soundness relative to the signed contract: an action lacking a formal proof of admissibility is blocked in enforce mode prior to reaching the target database, file system, or payment rail. Credentials remain inside the MITR contour rather than in the agent runtime.

MEASURED PROPERTY	EMPIRICAL VALUE (24-26 OCTOBER 2026 SNAPSHOT)	PRIMARY VERIFICATION SOURCE
Warm solver latency	1.08 ms median, 1.96 ms p95, 3.61 ms p99 across 1,000 decisions on 20 active contracts (4 CPU cores, 8 GB RAM, zero GPU).	`smt_latency_paper.jsonl`
Adversarial action suite	157 labeled scenarios (114 attacks, 43 routine operations): 0 missed attacks, 0 false refusals; 3 boundary assumptions documented.	`gate_attacks_runs.jsonl`
Cryptographic replay	5,094 receipts since 16 September 2026; epoch 33 anchored in Bitcoin block 973,024; standalone replay matched 5,079 of 5,079 records.	`asi5.ru/console/verify/`
Open engineering packages	Apache-2.0 distributions: `mcp-sentinel` 1.1.2, `iss-ledger-117` 1.1.1, `iss-agent-kit` 0.1.1, and standalone chain verifier.	`isslab.ai/mitr/start/`

4. Five-minute independent verification on a local workstation

Technical auditors and researchers can verify the public receipt anchors and test offline interception of destructive database calls in Python 3 without registration or network callbacks:

```
curl -O https://asi5.ru/console/verify/verify_chain_standalone.py.txt && mv verify_chain_standalone.py.txt
verify_chain_standalone.py
curl -O https://asi5.ru/console/verify/gate_anchors-public.jsonl
python3 verify_chain_standalone.py gate_anchors-public.jsonl  # exit code 0 confirms unbroken hash chain

mkdir wheels && cd wheels
curl -O https://isslab.ai/wp-content/uploads/iss-utils/mcp_sentinel-1.1.2-py3-none-any.whl
curl -O https://isslab.ai/wp-content/uploads/iss-utils/iss_ledger_117-1.1.1-py3-none-any.whl
cd .. && python3 -m venv .venv && . .venv/bin/activate && pip install --no-index --find-links ./wheels mcp-
sentinel
mcp-sentinel check --tool sql_query --args '{"query": "DROP TABLE users;"}'  # BLOCK, exit code 2
```

5. Bibliographic record and open-access editions

Citation: Fishchuk, Aleksandr. *Proven Action. Power and Control over AI Agents: Strategy 2027–2035* (original Russian edition: *Dokazannoe deystvie. Vlast i kontrol nad II-agentami: strategiya 2027–2035*). Moscow: Autonomous Non-Profit Organization “Scientific Research Institute of System Synthesis”, October 2026, 856 pages. Open-access Russian editions at `https://isslab.ai/book/`: 35-page architectural memorandum (`dokazannoe-deystvie-memorandum.pdf`) and 60-page introductory excerpt (`dokazannoe-deystvie-fragment.pdf`).

Correspondence, verification protocol, and enterprise pilots: English chapter summaries, claim registries, and SHA-256 checksums are maintained at `https://isslab.ai/book/` and `https://isslab.ai/book/verification-protocol/`.

Aleksandr Fishchuk
ORCID: 0009-0000-6549-6113
af@isslab.ai · @afishchuk