UNIVERSITY OF TARTU

SoK: Decentralized Agentic AI Security Agents in Smart Infrastructures

Olumba Chima Anya

Mubashar Iqbal · Raimundas Matulevičius



Funded by the European Union Horizon Europe programme · Grant Agreement No. 101217315

Introduction: Three Constraints on Smart Infrastructure Security

Millisecond timing

- Grid isolation 4–16 ms
- ICS safety 10–100 ms
- V2X collision 100–500 ms

Physical consequences

- Regional blackouts
- Damaged plant, injured staff
- Hazardous release

Scale and heterogeneity

- Thousands of nodes
- PLCs, RTUs, IEDs, sensors
- Unequal compute and protocols

Introduction: Why Centralized Security Falls Short



Single point of failure

One compromised hub blinds the entire defence at once.

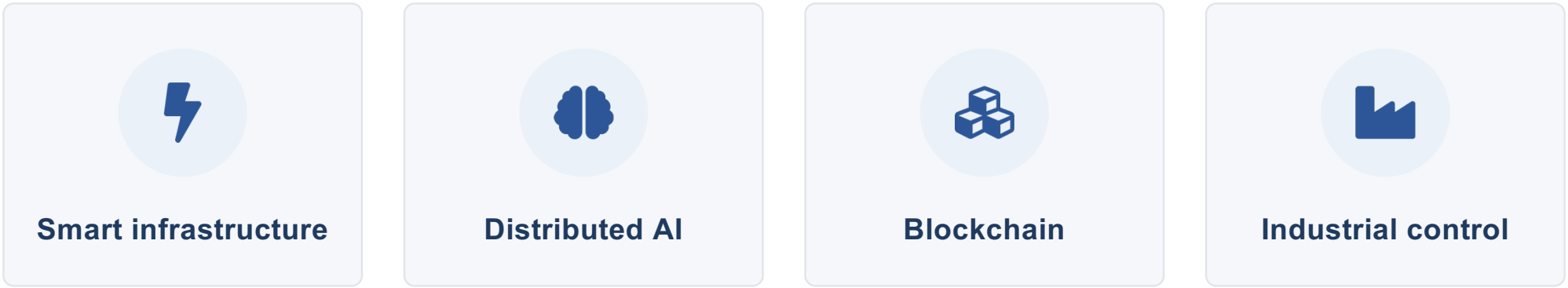


Round-trip latency

Report, decide, command cannot fit the protection window.

The field's response. Distribute the defence itself, and coordinate the agents without a central authority.

Motivation: Why a Systematization Is Needed



Four communities, publishing separately

How are decentralized coordination, agentic AI, and cybersecurity combined in smart infrastructures, and what purposes do they serve?

Research Questions

RQ1 What is the current state of decentralized agentic AI security in smart infrastructures?

RQ2 Which architectural patterns and technological components recur across the studies?

RQ3 Which capability gaps and contradictions remain in current approaches?

Scope: Three Definitions That Govern the Review

Security agent	Blockchain	Decentralization
Perceive	Audit trail	Centralized
Reason	Consensus	Distributed
Act	Smart-contract policy	Decentralized

Only the third arrangement counts as true decentralization in this review.

Contributions

Three-dimensional taxonomy

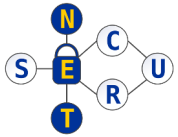
Compares architectures along Intelligence, Action, and Coordination.
Answers RQ2

Quantified capability coverage

Measures what the field has delivered, and names six persistent gaps.
Answers RQ1 and RQ3

Contradiction analysis

Tests claimed properties against the properties actually delivered.
Answers RQ3



Method: Systematic Literature Review

Kitchenham guidelines

govern how the review is conducted

PRISMA 2020 statement

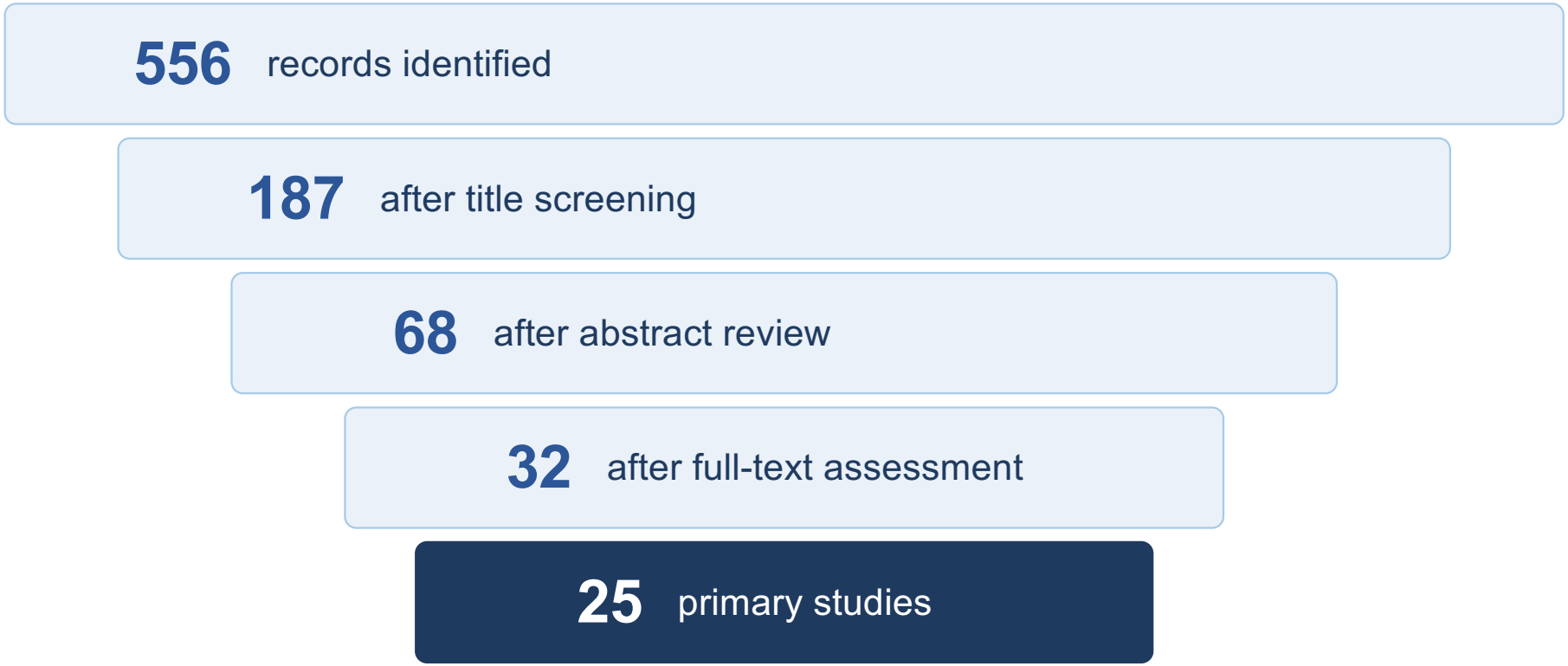
governs how every stage is reported

Peer-reviewed, 2020–2026. ACM Digital Library · IEEE Xplore · SpringerLink · ScienceDirect

A study had to satisfy all three search areas

Domain | **Security function** | **Architecture**

Method: Study Selection



Results: Domains and Technological Components

ANSWERING RQ1 + RQ2

Application domains

Smart grid / power	44%
Industrial IoT	32%
Intelligent transport	24%

AI techniques

Deep neural networks	56%
Federated learning	24%
RL / DRL	16%
Traditional ML	12%
Rule-based	8%

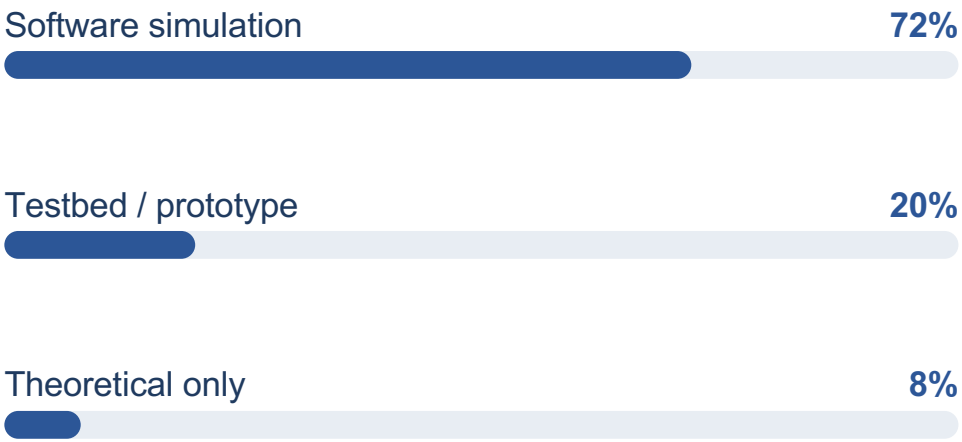
Blockchain platforms

Ethereum	44%
Custom / unspecified	36%
Hyperledger Fabric	20%

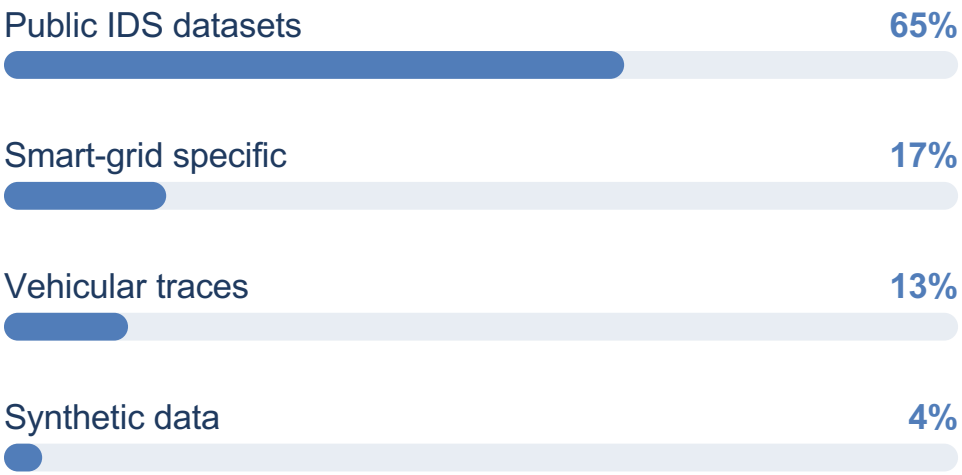
Results: How Rigorously Was It Evaluated?

ANSWERING RQ1

Evaluation environment (n = 25)

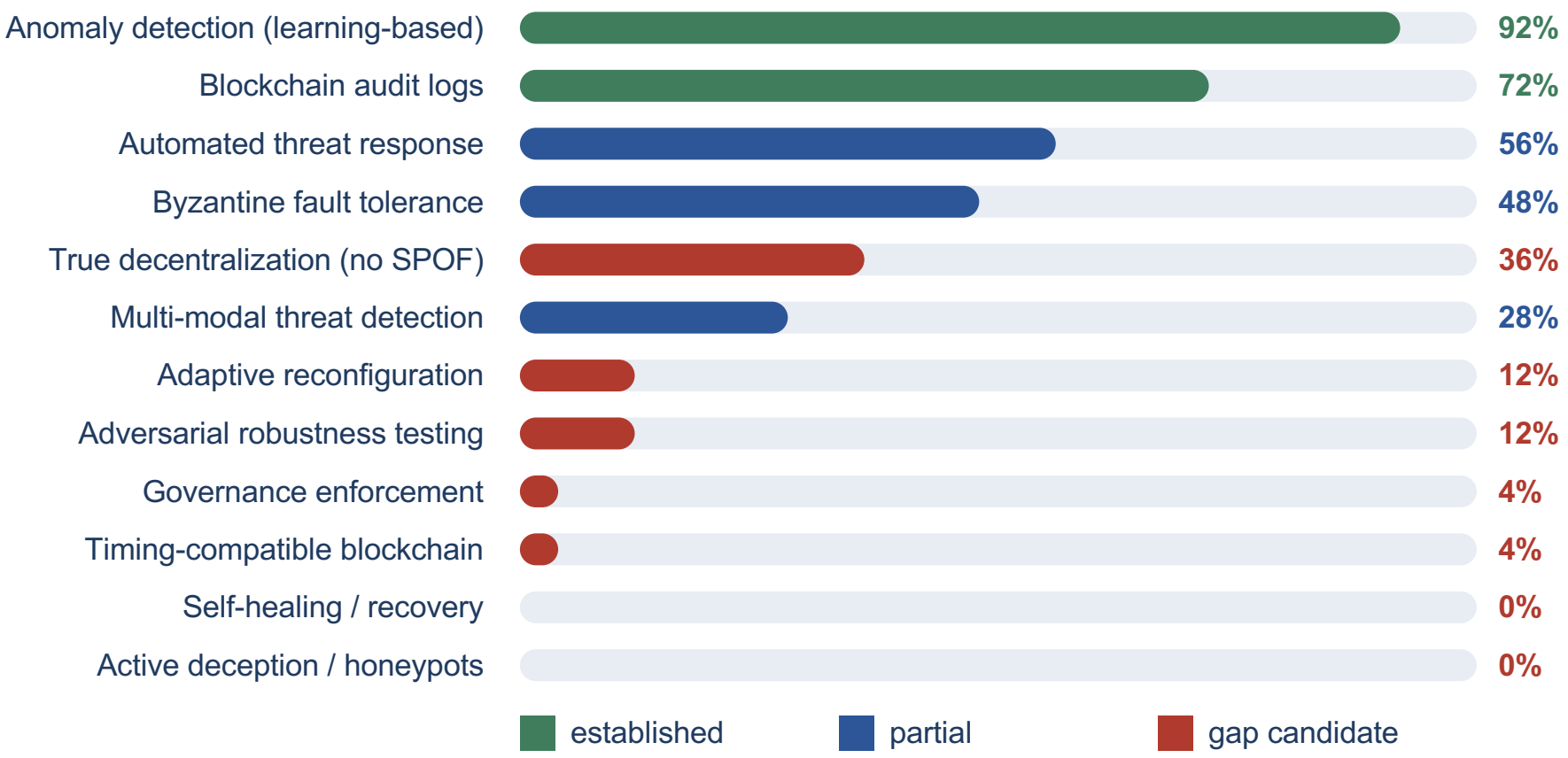


Datasets among 23 empirical studies



Results: Capability Coverage Across the 25 Studies

ANSWERING RQ1



Analysis: The Three-Dimensional Taxonomy

ANSWERING RQ2

INTELLIGENCE how agents perceive and learn	ACTION which operations agents execute	COORDINATION how agents agree and synchronize
Active deception		
Adversarial robustness	Recovery and healing	
Federated learning	Adaptive reconfiguration	Peer-to-peer
Adaptive (RL / DRL)	Active response	Blockchain
Learning-based (DNN)	Passive response	Hierarchical
Reactive (rule-based)	Detect and alert	Centralized

Each dimension runs from lower to higher capability and cost, not from worse to better for every domain.

Analysis: Four Recurring Architectural Patterns

ANSWERING RQ2

P1 **Detection plus audit logs**

Accountability without any active intervention.

P2 **Federated with central aggregation**

Better accuracy at the price of a central server.

P3 **Active response, no recovery**

Nodes are isolated, then stay offline.

P4 **Blockchain, no timing analysis**

Ledger adopted without checking the latency budget.

Results: Six Persistent Capability Gaps

ANSWERING RQ3

A capability counts as a gap when it is absent, or present in fewer than half the studies.

G1

Active deception

0% coverage

G2

Self-healing recovery

0% coverage

G3

Governance enforcement

4% coverage

G4

Timing-compatible blockchain

4% coverage

G5

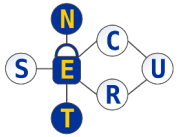
True decentralization

36% coverage

G6

Adversarial robustness

12% coverage



Results: The Integration Paradox

ANSWERING RQ3

11

studies address
none of the six gaps

14

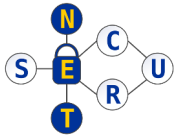
studies address
exactly one

0

studies address
two or more

No reviewed architecture combines even two gap capabilities under realistic timing, safety, and distribution constraints.

Bounded by how this corpus was defined and coded.



Results: Timing Budgets versus Consensus Latency

ANSWERING RQ3

Domains require

Smart grid fault isolation	4–16 ms
Industrial control safety	10–100 ms
Vehicular collision avoidance	100–500 ms

Blockchain delivers

Ethereum block confirmation	12–15 s
Hyperledger Fabric commit	0.5–2 s

Up to 3,750× over the grid protection budget

Hard real-time → local to the agent

Fault isolation, node quarantine, safety interlocks. No consensus in the loop.

Soft real-time → over the ledger

Audit logging, forensics, policy distribution. Seconds of latency are acceptable.

Results: Claimed versus Delivered Properties

ANSWERING RQ3

“Decentralized”

56%

of studies claim it



yet retain a central aggregation server or policy hub

“Real-time”

44%

of studies claim it



yet run on ledgers whose latency exceeds the domain budget

“Autonomous”

16%

of studies claim it



yet require central permission for high-impact actions

The field has no agreed criteria for validating any of these three claims.

Limitations and Future Work

Limitations

- Peer-reviewed sources only
- Some classification required judgement
- Percentages may shift at the margins
- No operational deployment in the corpus

Future work

Testbed validation
check whether the gaps persist under operational conditions

Gap-pair prototypes
target two or more gaps inside one architecture

Domain adaptation
grid, ICS, and transport variants built with operators

Conclusion

A structured taxonomy

classifies designs along Intelligence, Action, and Coordination

Six quantified gaps

so completeness can be assessed instead of assumed

Three contradictions

so decentralization, real-time, and autonomy claims can be tested

Thank you · Questions?

olumba.chima.anya@ut.ee · ORCID 0009-0005-8092-2535

Funded by the European Union Horizon Europe programme · Grant No. 101217315