

Research Paper Proposal

ASBOS

A Risk-Aware Multi-Agent AI Framework for Autonomous Private Security Operations

Prepared by

K.M. Fazle Rabbi

Department of Computer Science & Engineering
St. Cloud State University

Research Advisor

Dr. Indira Dutta

October 2026

1. Topic Sentence

This research investigates how a risk-aware multi-agent AI system can monitor, coordinate, and automate routine private security operations while keeping human supervisors responsible for high-risk decisions.

2. Introduction

Private security companies depend on many operational processes that happen continuously behind the scenes. These include scheduling security officers, monitoring active shifts, tracking guard locations, detecting late arrivals and no-shows, verifying patrols, managing incidents, finding replacement officers, checking overtime, maintaining compliance records, and communicating with supervisors and clients.

Most security management systems help store and organize this information, but the actual coordination and decision-making are still heavily dependent on human supervisors, dispatchers, and managers.

For example, when a security officer does not arrive for a scheduled shift, a supervisor may need to notice the problem, contact the officer, determine whether the site is uncovered, identify qualified replacements, check availability and overtime, assign another officer, update the schedule, and document the change.

Similar manual processes occur when patrols are missed, guards leave assigned locations, incidents occur, or coverage suddenly becomes unavailable.

Recent advances in autonomous AI agents and multi-agent systems create an opportunity to investigate whether some of these operational responsibilities can be automated. Multi-agent systems can divide complex problems among specialized agents that communicate and coordinate with each other [4, 7].

Rather than allowing one AI system to control an entire organization, specialized AI agents could be responsible for different operational areas. These agents could communicate with one another, monitor real-time events, recommend actions, perform routine actions, and escalate higher-risk situations to human supervisors.

This research proposes **ASBOS**, a risk-aware multi-agent AI framework designed to coordinate routine private security operations while preserving meaningful human oversight.

3. Problem Statement

Private security operations generate continuous information from multiple sources, including:

- employee schedules,
- GPS and geofence information,
- clock-in and clock-out records,
- patrol checkpoints,
- incident reports,
- employee qualifications,
- site requirements,
- overtime records,
- supervisor communications, and
- client requirements.

Although this information may already exist digitally, responding to operational problems often requires significant manual monitoring and coordination.

Traditional automation can respond to predefined rules, but real security operations are dynamic. One event may affect several parts of an organization at the same time.

For example, if a security officer does not arrive for a shift, a system may need to:

1. detect the absence,
2. determine whether the location is uncovered,
3. contact the assigned officer,
4. identify available replacement officers,
5. verify qualifications,
6. check overtime limits,
7. consider travel distance,

8. notify a supervisor,
9. update the schedule,
10. document the operational change, and
11. possibly notify the client.

This is not a single isolated task. It is a sequence of related decisions involving multiple operational systems and different levels of risk.

The main problem addressed in this research is how multiple specialized AI agents can coordinate these decisions while determining which actions can safely be automated and which decisions should require human approval.

4. Research Gap

Research already exists in multi-agent systems, autonomous AI agents, automated scheduling, human-automation collaboration, and security resource allocation.

Research in security games has also examined computational approaches for allocating security resources and creating patrol schedules under operational constraints [3].

Recent work has explored large-language-model-based multi-agent systems where specialized agents communicate and collaborate on complex tasks [4, 1].

However, these areas are commonly examined separately.

There is an opportunity to investigate how multi-agent coordination, real-time operational information, workforce management, physical security operations, and risk-based human oversight can be combined into one operational framework specifically for private security organizations.

ASBOS does not attempt to eliminate human involvement from security operations.

Instead, it investigates how routine operational coordination can be delegated to autonomous AI agents while keeping humans responsible for decisions where mistakes could create significant safety, employment, legal, financial, or client consequences.

5. Research Objective

The primary objective of this research is to design and evaluate a multi-agent AI framework capable of coordinating routine private security operations while adapting its level of autonomy according to operational risk.

The research will investigate whether specialized AI agents can reduce manual operational workload while maintaining appropriate human control.

6. Main Research Question

How can a risk-aware multi-agent AI system automate routine private security operations using real-time operational data while preserving human oversight for high-risk decisions?

Supporting Research Questions

RQ1:

Can specialized AI agents coordinate security operations more effectively than a single-agent or traditional rule-based automation system?

RQ2:

How should the level of AI autonomy change according to the operational risk of a decision?

RQ3:

How much human operational workload can be reduced without increasing incorrect or unsafe autonomous actions?

RQ4:

How effectively can a multi-agent system respond to interconnected events involving guard tracking, scheduling, patrol compliance, and incident management?

7. Proposed ASBOS Architecture

ASBOS will use several specialized AI agents. Each agent will focus on a specific operational area while communicating with a coordinating agent.

7.1 Field Operations Agent

The Field Operations Agent will monitor:

- guard GPS and geofence status,
- clock-ins and clock-outs,
- patrol checkpoints,
- missed patrols,
- unexpected location changes,
- shift activity, and
- current guard status.

7.2 Workforce Agent

The Workforce Agent will manage:

- employee schedules,
- late arrivals,
- no-shows,
- replacement officer availability,
- overtime,
- employee qualifications,
- scheduling conflicts, and
- shift coverage.

7.3 Incident and Dispatch Agent

The Incident and Dispatch Agent will manage:

- incident reports,
- alarms,

- requests for assistance,
- supervisor notifications,
- missing incident information,
- follow-up activities,
- dispatch coordination, and
- escalation.

7.4 Compliance and Business Operations Agent

This agent will evaluate:

- employee licenses,
- training requirements,
- site-specific requirements,
- working-hour restrictions,
- timesheet exceptions,
- overtime policies, and
- organizational policies.

7.5 Supervisor and Coordinator Agent

The Supervisor and Coordinator Agent will receive information from the specialized agents and evaluate the overall operational situation.

It will determine whether:

- an event requires no action,
- an automated action is appropriate,
- a recommendation should be sent to a supervisor,
- human approval is required, or
- an event requires immediate human escalation.

The core ASBOS workflow can be represented as:

Monitor → Detect → Reason → Coordinate → Act or Escalate

8. Risk-Based Levels of Autonomy

Levels of automation and human interaction have been studied extensively in human-automation research [5]. ASBOS adapts this concept for private security operations.

Level 0 — Observation

The system monitors operations, identifies events, and records information. No operational decisions are made automatically.

Level 1 — Recommendation

The AI detects a problem, analyzes the situation, and recommends an action. A human makes the final decision.

Level 2 — Approval Required

The AI determines and prepares an action, but a human supervisor must approve it before execution.

Level 3 — Routine Autonomy

The AI can automatically perform predefined low-risk operational actions. The action is logged and available for human review.

Level 4 — Coordinated Autonomy

Multiple specialized agents may collaborate to resolve an operational event automatically when the event remains within established policy and risk boundaries.

High-risk decisions will remain under human control.

9. Experimental Scenarios

The initial ASBOS prototype will focus on four realistic operational scenarios.

9.1 Scenario 1: Late Arrival or No-Show

A security officer does not arrive for a scheduled shift.

ASBOS must:

- detect the problem,
- determine whether the post is uncovered,
- attempt to contact the officer,
- evaluate available replacements,
- verify qualifications,
- check overtime,
- consider travel distance,
- recommend or assign coverage, and
- determine whether supervisor approval is required.

9.2 Scenario 2: Guard Leaves Assigned Geofence

An on-duty security officer unexpectedly leaves an assigned site or geofenced area.

The system must determine whether the event represents:

- normal activity,
- an authorized exception,
- a temporary operational issue,
- a possible coverage problem, or
- an event requiring supervisor attention.

The system should not automatically interpret a location anomaly as employee misconduct.

9.3 Scenario 3: Missed Patrol Checkpoint

A required patrol checkpoint is not completed within the expected time.

ASBOS may evaluate:

- previous checkpoint activity,
- time since last activity,
- current location,
- site requirements,
- active incidents, and
- officer status.

It will then determine whether to continue monitoring, contact the officer, recommend an action, or escalate the event.

9.4 Scenario 4: Incident With Coverage Impact

An incident requires an officer to leave a normal assignment or requires additional personnel.

Multiple ASBOS agents must coordinate:

- incident response,
- staffing,
- replacement coverage,
- supervisor notification,
- documentation, and
- operational follow-up.

10. Research Methodology

A prototype of ASBOS will be developed using an event-driven architecture.

The initial study will use simulated or de-identified operational information rather than requiring live employee tracking data.

The simulated environment may generate:

- shift start and end events,
- GPS and geofence events,

- guard availability,
- clock-in and clock-out events,
- patrol checkpoints,
- incidents,
- qualifications,
- overtime information,
- site requirements, and
- supervisor decisions.

Three approaches will be compared.

10.1 Approach A: Rule-Based System

A traditional automation system will use predetermined rules to respond to events.

10.2 Approach B: Single AI Agent

One AI agent will receive all available operational information and determine an appropriate response.

10.3 Approach C: ASBOS Multi-Agent System

Specialized AI agents will analyze their respective operational areas and coordinate through the ASBOS architecture.

All three systems will receive equivalent test scenarios.

11. Evaluation Metrics

Potential metrics include:

- successful event-resolution rate,
- average response time,
- incorrect action rate,

- unsafe action rate,
- correct escalation rate,
- unnecessary escalation rate,
- number of human interventions,
- percentage of operational steps automated,
- decision consistency,
- policy violations,
- decision auditability, and
- estimated reduction in human operational workload.

The objective is not simply to maximize automation.

A system that performs more autonomous actions but produces unsafe or incorrect decisions will not be considered superior.

12. Research and Development Cost

ASBOS is intended to be tested as an operational prototype rather than only as a theoretical multi-agent simulation. Therefore, the research requires several third-party services for communication, location services, AI processing, data storage, event delivery, and infrastructure.

The following budget represents the estimated cost of designing and developing the initial ASBOS research prototype.

Resource	Purpose	Estimated Cost
AI / LLM API usage	Development and debugging of specialized ASBOS agents, coordinator reasoning, classification, and decision generation	\$250
Cloud hosting, database, storage, and background services	Application hosting, PostgreSQL database, event processing, object storage, queues, and development infrastructure	\$150
Telnyx communication services	Development testing for SMS, telephone numbers, voice communication, alerts, and communication webhooks	\$100
Resend email infrastructure	Transactional email, operational notifications, email workflow development, and delivery testing	\$80
Domains, DNS, and test communication configuration	Research domains, subdomains, email authentication, DNS configuration, and isolated test environments	\$40
Maps, geocoding, and location services	Geofence testing, simulated guard locations, distance calculation, and location-aware operational decisions	\$75
Logging and monitoring	System logs, agent activity tracking, error monitoring, and audit records	\$50
Synthetic data and automation tools	Generation of simulated guards, shifts, incidents, patrols, schedules, and operational events	\$50
Development contingency	Unexpected API usage, service fees, additional infrastructure, or implementation requirements	\$105
Estimated Research and Development Total		\$900

Table 1: Estimated ASBOS research and development budget.

The development budget is intentionally based on paid or usage-based services where necessary. Free tiers may be used during early development, but the research budget does not assume that all experiments can remain within free service limits.

13. Testing and Evaluation Cost

The experimental testing phase is expected to require greater third-party usage than normal application development.

ASBOS will be evaluated through repeated operational scenarios involving communications, guard tracking, agent coordination, scheduling events, incident escalation, and human-oversight decisions. Some experiments may be executed hundreds or thousands of times to compare the rule-based system, single-agent system, and ASBOS multi-agent architecture.

Mass email, SMS, voice, and other communication tests are included because communication is an important part of real private security operations.

Testing Resource	Purpose	Estimated Cost
Telnyx SMS and voice testing	Large-scale testing of guard alerts, supervisor notifications, no-show workflows, escalation messages, voice events, and communication failures	\$250
Resend mass email testing	Testing high-volume operational emails, supervisor notifications, client communications, incident notifications, automated reports, and email delivery events	\$200
AI / LLM experimental runs	Repeated execution of single-agent and multi-agent scenarios under different operational conditions	\$250
Cloud compute, database, and infrastructure	Running repeated experiments, storing event history, processing background jobs, and maintaining testing environments	\$150
Location and geofence testing	Repeated GPS events, distance calculations, movement simulations, and geofence-entry and exit scenarios	\$75
Load testing, queues, webhooks, and event simulation	Testing concurrent security events, communication failures, event queues, retries, and multi-agent coordination	\$75
Test phone numbers, email identities, and test domains	Maintaining isolated communication identities and endpoints for experimental testing	\$50
Data storage and experimental analysis	Storage and export of experiment logs, agent decisions, metrics, and evaluation results	\$25
Testing contingency	Additional usage charges and unexpected experimental requirements	\$25
Estimated Testing and Evaluation Total		\$1,100

Table 2: Estimated ASBOS testing and evaluation budget.

14. Estimated Total Research Budget

The estimated overall budget for the ASBOS research project is approximately \$2,000.

Budget Category	Estimated Cost
Research and prototype development	\$900
Experimental testing and evaluation	\$1,100
Total Estimated Research Budget	\$2,000

Table 3: Overall estimated ASBOS research budget.

The \$2,000 figure should be considered a research budget estimate rather than a fixed expenditure. Actual spending will depend on the number of experiments, communication volume, AI model selection, cloud usage, and the number of operational scenarios tested.

Some services may provide free tiers or university credits; however, this proposal intentionally includes sufficient funding for paid testing so that experimental design is not restricted by free-tier limits.

The budget is also intended to support realistic testing of third-party dependencies. For example, ASBOS may need to evaluate how the system behaves when an SMS fails, an email is delayed, a webhook is retried, a telephone call is unanswered, a GPS event is missing, or several operational events occur simultaneously. These tests are important because an autonomous operational system must be evaluated under both normal and failure conditions.

15. Expected Contribution

The primary contribution of this research is a framework for determining how operational autonomy can be distributed between specialized AI agents and human supervisors in private security operations.

The study is expected to contribute:

1. a domain-specific multi-agent architecture for private security operations;
2. a risk-based model for determining AI autonomy;
3. an event-driven method for coordinating operational agents;
4. realistic private-security testing scenarios;

5. a framework for human escalation and oversight; and
6. experimental evidence comparing multi-agent, single-agent, and rule-based approaches.

16. Ethical, Privacy, and Safety Considerations

Guard tracking and workforce monitoring create important privacy and employment concerns.

Location information should only be collected when operationally necessary and should not automatically be interpreted as evidence of employee misconduct.

AI-generated decisions should be logged so supervisors can determine what information was considered and why an action was recommended or executed.

Risk management and clearly defined human responsibilities are important when AI systems are used in operational environments [6]. Recent research on agentic systems also emphasizes the importance of oversight throughout autonomous workflows [2].

The research prototype will not allow AI agents to independently make high-impact decisions such as:

- employee termination,
- disciplinary action,
- hiring decisions,
- law-enforcement actions,
- accusations of misconduct, or
- other significant employment decisions.

The study will focus on operational coordination rather than automated employee punishment.

17. Scope and Limitations

The initial study will not attempt to automate every function of a private security company.

The primary scope will focus on:

Monitor → Detect → Reason → Coordinate → Act/Escalate

Areas outside the initial scope include:

- payroll processing,
- recruiting,
- sales,
- invoicing,
- contract management,
- advanced video surveillance analysis,
- autonomous disciplinary decisions, and
- complete company administration.

18. Future Work

Future versions of ASBOS could investigate:

- automated client communication,
- predictive staffing,
- security resource optimization,
- automated report review,
- training and certification monitoring,
- equipment and asset management,
- operational risk forecasting,
- integration with video analytics,
- automated billing reconciliation, and
- larger-scale autonomous company coordination.

A longer-term research direction could investigate whether increasingly capable autonomous systems allow a significantly smaller management team to supervise larger private security operations.

19. Conclusion

Private security operations involve continuous monitoring, communication, coordination, and decision-making.

Although many security companies use digital systems for scheduling, reporting, tracking, and workforce management, those systems generally depend on humans to identify operational problems and coordinate responses.

ASBOS proposes a different model in which several specialized AI agents continuously monitor different operational areas and collaborate when problems occur.

The purpose of ASBOS is not to eliminate human supervisors. Instead, this research investigates where autonomy is useful, where human approval remains necessary, and how multiple AI agents can work together while remaining within clearly defined operational and risk boundaries.

By comparing ASBOS against traditional rule-based automation and a single-agent AI approach, the research aims to produce measurable evidence about the benefits, risks, and limitations of multi-agent artificial intelligence for private security operations.

References

- [1] Bajoria, S., Ranjan, S., M, A., Suri, V., Hassija, V., Chamola, V., & Hussain, A. (2026). *From language models to agentic AI: A survey of autonomous, action-enabled, and collaborative LLM agents*. Cognitive Computation, 18, 103. <https://doi.org/10.1007/s12559-026-10619-1>
- [2] Dhanorkar, S., Passi, S., & Vorvoreanu, M. (2026). *Human oversight of agentic systems in practice: Examining the oversight work, challenges, and heuristics of developers using software agents*. arXiv preprint arXiv:2606.05391. <https://doi.org/10.48550/arXiv.2606.05391>
- [3] Jain, M., Kardes, E., Kiekintveld, C., Ordonez, F., & Tambe, M. (2010). Security games with arbitrary schedules: A branch and price approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, 24(1), 792–797. <https://doi.org/10.1609/aaai.v24i1.7611>
- [4] Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y. (2024). A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges. *Vicinagearth*, 1, 9. <https://doi.org/10.1007/s44336-024-00009-2>
- [5] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3), 286–297. <https://doi.org/10.1109/3468.844354>
- [6] Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. NIST AI 100-1. <https://doi.org/10.6028/NIST.AI.100-1>
- [7] Tran, K.-T., Dao, D., Nguyen, M.-D., Pham, Q.-V., O’Sullivan, B., & Nguyen, H. D. (2025). *Multi-agent collaboration mechanisms: A survey of LLMs*. arXiv preprint arXiv:2501.06322.