APPLICATION OF LARGE LANGUAGE MODELS FOR AUTOMATED ANALYSIS OF CYBERSECURITY INCIDENTS

Khusenov Murodjon Zokhirovich

Associate Professor, Department of Artificial Intelligence and Information Security, Bukhara State University, Bukhara, Uzbekistan

E-mail: m.z.xusenov@buxdu.uz ORCID: 0000-0002-1533-3102

Abstract. Security operations centres (SOCs) face a growing volume of alerts, analyst fatigue and a shortage of qualified specialists, which makes the timely and high-quality analysis of cybersecurity incidents increasingly difficult. Large language models (LLMs), owing to their ability to interpret unstructured texts — logs, alerts, threat intelligence reports — are considered in the recent literature as a promising means of automating this analysis. This article systematizes the documented capabilities of LLMs across the phases of the incident response lifecycle defined by NIST, proposes a five-layer reference architecture for the responsible integration of LLMs into incident analysis processes (data ingestion, knowledge and retrieval, model and orchestration, assurance, audit and compliance), and constructs a structured map of the principal risks of such integration — hallucinations, prompt injection, leakage of sensitive data, adversarial evasion, over-reliance and integration with legacy infrastructure — together with mitigation measures grounded in current guidance. The conclusion is substantiated that, at the present level of technology, LLMs should augment rather than replace the analyst, with mandatory human approval of consequential response actions. The results are oriented towards SOCs of organizations and universities, including in the context of the implementation of the cybersecurity strategy of the Republic of Uzbekistan for 2026–2030.

Keywords: *large language models; cybersecurity incident analysis; security operations centre; incident response; retrieval-augmented generation; prompt injection; human-in-the-loop*

1. Introduction

The analysis of cybersecurity incidents remains one of the most labour-intensive processes of operational security. Surveys of the application of artificial intelligence in security operations centres record a stable triad of problems: the volume of alerts generated by monitoring systems exceeds the processing capacity of analysts; routine triage produces fatigue and the risk of missing significant events; and the shortage of qualified specialists limits the scaling of manual analysis [1]. The normative basis of incident handling is formed by the recommendations of the National Institute of Standards and Technology: the current revision of Special Publication 800-61 integrates incident response into the overall cybersecurity risk management of the organization and structures it through the functions of the Cybersecurity Framework, from preparation to detection, response, recovery and post-incident improvement [2].

The emergence of large language models has changed the technological context of this task. The transformer architecture [3] and the demonstration that language models scale to few-shot learning of new tasks [4] produced systems capable of summarizing, classifying and generating professional text from heterogeneous inputs. Precisely these properties correspond to the texture of incident analysis, which operates on unstructured logs, alert descriptions, threat intelligence reports and correspondence. Recent surveys document the application of LLMs to log analysis and summarization, alert triage and prioritization, enrichment of events with threat intelligence context, mapping of observed behaviour to adversary technique taxonomies, and the drafting of incident reports [1, 5, 6].

At the same time, the same surveys emphasize that LLM integration introduces new risks — from hallucinated conclusions and prompt injection through processed content to the leakage of sensitive telemetry — and that evaluation methodologies for such systems are still immature [5, 6, 8]. For the Republic of Uzbekistan the question has practical urgency: the national cybersecurity strategy for 2026–2030 explicitly prioritizes the development and implementation of cybersecurity technologies based

on artificial intelligence and requires the creation of cybersecurity units in state bodies [7]. The aim of this study is to substantiate an approach to the application of LLMs for the automated analysis of cybersecurity incidents that realizes their documented capabilities while controlling the associated risks. The objectives are: (i) to systematize the functions of LLMs across the phases of the incident response lifecycle; (ii) to propose a reference architecture of integration; (iii) to construct a map of risks and mitigation measures.

2. Materials and Methods

The study is based on the analysis of recent surveys and primary publications on the application of LLMs in security operations, including a PRISMA-guided systematic review of LLM- and agent-based security automation [5], a survey of LLM applications, vulnerabilities and defence techniques in cybersecurity [6], and a survey of AI in security operations centres [1], complemented by normative and methodological documents: NIST Special Publication 800-61 Revision 3 on incident response [2], the MITRE ATT&CK knowledge base of adversary tactics and techniques [9], and the OWASP Top 10 for Large Language Model Applications as a catalogue of LLM-specific risks [8].

Three methods were applied. First, the documented functions of LLMs were mapped onto the phases of the incident response lifecycle in the structure of SP 800-61, which makes it possible to see where automation is mature and where it remains experimental. Second, a reference architecture was synthesized from the engineering practices recurring in the analysed corpus, guided by three design principles: grounding of model outputs in verifiable organizational data; least privilege in the model's access to systems and actions; and auditability of every automated step. Third, the risks identified in the surveys were aligned with the OWASP LLM risk catalogue [8] and with mitigation measures documented in the literature. The study is a structured review with architectural synthesis; it does not report original experimental measurements, and

the empirical evaluation of the proposed architecture is defined as the next stage of the research.

3. Results

3.1. Functional taxonomy across the incident response lifecycle. In the preparation phase, LLMs are applied to the authoring and updating of playbooks and to the explanation of detection rules and configurations in natural language, lowering the entry barrier for junior analysts [1, 5]. In the detection and analysis phase — the most saturated with documented applications — LLMs summarize voluminous and heterogeneous logs, triage and prioritize alerts, enrich events with context from threat intelligence sources, map observed behaviour onto the MITRE ATT&CK taxonomy [9], and support timeline reconstruction in digital forensics; the systematic review describes, in particular, generative frameworks for forensic timeline analysis of incident events [5]. In the containment, eradication and recovery phase, LLMs draft response actions and parameterize playbooks of orchestration (SOAR) platforms; here the literature is unanimous that consequential actions — isolation of hosts, blocking of accounts — must pass through human approval [1, 5]. In the post-incident phase, LLMs generate structured incident reports, including summary tables of identified adversary techniques, and support the extraction of lessons learned for the improvement of defences [5].

3.2. Reference architecture. The proposed architecture comprises five layers. The data ingestion layer normalizes telemetry from SIEM and EDR systems, network sensors and service logs. The knowledge and retrieval layer implements retrieval-augmented generation over the organization’s own sources — asset inventory, network documentation, internal policies, accumulated incident reports and threat intelligence feeds — so that model conclusions are grounded in verifiable organizational facts rather than in parametric memory alone. The model and orchestration layer hosts the language models and agent logic: task decomposition, tool calling for queries to security systems, and structured output schemas that make responses machine-

checkable. The assurance layer implements human-in-the-loop gates: automatically executed steps are limited to read-only enrichment, whereas any state-changing action requires analyst confirmation; the same layer applies validation rules to model outputs. The audit and compliance layer logs every prompt, retrieved document, model response and approval decision, which is necessary both for incident investigation quality control and for accountability. For environments processing sensitive telemetry, the architecture provides for locally hosted or organization-controlled models, so that incident data do not leave the protected perimeter [5, 6].

Table 1. Principal risks of LLM integration into incident analysis and mitigation measures

Risk	Manifestation in incident analysis	Mitigation measures
Hallucinations	Fabricated indicators, incorrect attribution of techniques, plausible but wrong conclusions in reports [5]	Retrieval grounding in organizational data; structured output validation; human review of conclusions before action
Prompt injection	Malicious instructions embedded in analysed logs, phishing samples or threat intelligence content [8]	Treating all analysed content as untrusted data; input sanitization; privilege separation between analysis and action; OWASP LLM controls [8]
Sensitive data leakage	Incident telemetry, credentials or personal data exposed to external model providers [5]	Locally hosted or organization-controlled models; masking and minimization of data in prompts; data processing agreements
Adversarial evasion	Attackers crafting inputs that mislead model-based triage and classification [6]	Adversarial testing and red teaming of the pipeline; ensemble checks; monitoring of model performance drift
Over-reliance of analysts	Erosion of analyst skills and uncritical acceptance of model output	Training that includes model failure modes; mandatory explanation of conclusions; periodic manual exercises
Integration with legacy infrastructure	Incompatibility with existing SIEM/SOAR processes and data formats [5]	API gateways and standardized schemas; phased rollout starting from low-risk enrichment functions

3.3. Risk map. Table 1 systematizes the principal risks identified in the analysed corpus and aligned with the OWASP catalogue of LLM application risks [8], together with mitigation measures. The structure of the table reflects the design principles of the architecture: grounding counters hallucinations, privilege separation and input handling counter prompt injection, deployment control counters data leakage, and the assurance and audit layers operationalize the remaining measures.

4. Discussion

The juxtaposition of the functional taxonomy and the risk map leads to the central conclusion: at the present level of technology, the rational mode of applying LLMs in incident analysis is augmentation of the analyst, not autonomous response. The documented strengths of the models — summarization, triage, enrichment, report generation — lie in the phases where the cost of an error is correctable, whereas state-changing response actions must remain behind human approval gates; this position is consistent across the analysed surveys [1, 5, 6]. The proposed five-layer architecture formalizes this mode and makes it auditable.

A second conclusion concerns evaluation. The surveys agree that benchmarks for LLM-based security automation are immature: reported results are often obtained on heterogeneous private datasets, which complicates comparison and procurement decisions [5, 6]. For organizations this argues for pilot deployments with locally defined metrics — triage time, escalation precision, report completeness — before any expansion of model authority. For the research community it defines a priority: open, realistic incident analysis benchmarks.

For the Republic of Uzbekistan the results have a direct projection. The national cybersecurity strategy for 2026–2030 simultaneously prioritizes AI-based protection technologies and obliges state bodies to create cybersecurity units, with multi-factor authentication requirements for state information systems and a register of outsourcing service providers [7]; the Law “On Cybersecurity” establishes the general legal framework of these processes [10]. Under personnel shortage, LLM-based

augmentation of incident analysis is one of the few realistic ways to raise the productivity of small security teams, including the emerging security functions of universities; the architecture proposed here — in particular, locally hosted models and audit logging — is compatible with the data protection requirements of such organizations. The limitation of the study is its review-analytical character: the architecture is synthesized from the literature, and its empirical evaluation on the incident flow of a concrete organization, including a university SOC, is planned as the next stage of this work.

5. Conclusion

Large language models have formed a documented set of capabilities for the automation of cybersecurity incident analysis: summarization of logs and alerts, triage, enrichment with threat intelligence, mapping to adversary technique taxonomies, drafting of response steps and generation of structured reports. The contribution of this article is threefold: a taxonomy of these functions across the phases of the NIST incident response lifecycle; a five-layer reference architecture (data ingestion, knowledge and retrieval, model and orchestration, assurance, audit and compliance) realizing the principles of grounding, least privilege and auditability; and a structured map of risks and mitigations aligned with the OWASP catalogue of LLM application risks. The substantiated mode of application is augmentation of the analyst with mandatory human approval of consequential actions. Further research will be directed at the empirical evaluation of the architecture on real incident flows, the development of open benchmarks, and the adaptation of the approach to the conditions of university security operations in Uzbekistan.

References

1. Habibzadeh A. et al. Large Language Models in Security Operations Centers: A Survey // arXiv preprint arXiv:2509.10858. 2025.

2. Incident Response Recommendations and Considerations for Cybersecurity Risk Management: A CSF 2.0 Community Profile. NIST Special Publication 800-61, Revision 3. Gaithersburg: National Institute of Standards and Technology, 2025.
3. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. Attention Is All You Need // Advances in Neural Information Processing Systems 30 (NeurIPS 2017). 2017.
4. Brown T. B. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems 33 (NeurIPS 2020). 2020.
5. AI-Augmented SOC: A Survey of LLMs and Agents for Security Automation // Journal of Cybersecurity and Privacy (MDPI). 2025. Vol. 5, No. 4. Article 95.
6. Large Language Models in Cybersecurity: A Survey of Applications, Vulnerabilities, and Defense Techniques // AI (MDPI). 2025. Vol. 6, No. 9. Article 216.
7. Abdushukurov N. Uzbekistan adopts national cybersecurity strategy to 2030. 2026. [Online]. Available: https://www.nurilla.com/insights/uzbekistan-cybersecurity-strategy-2030.html
8. OWASP Top 10 for Large Language Model Applications. OWASP Foundation. [Online]. Available: https://owasp.org/www-project-top-10-for-large-language-model-applications/
9. MITRE ATT&CK: A Knowledge Base of Adversary Tactics and Techniques. The MITRE Corporation. [Online]. Available: https://attack.mitre.org
10. Law of the Republic of Uzbekistan “On Cybersecurity” No. LRU-764. April 15, 2022.