Generated by Scopus AI, Thu Dec 18 2025

Automating cyber range scenario generation with LLMs

State-of-the-Art in Automated Cyber Range Scenario Generation Using Large Language Models: Addressing Hallucinations, Context Drift, and Logical Consistency with Hybrid AI Approaches (2023–2025)

Quick Reference

Key Findings Table

Theme	Key Insights	Supporting Citations
Hybrid AI Integration	Neuro-symbolic AI, Knowledge Graphs, and Multi-Agent Systems are increasingly combined with LLMs to mitigate hallucinations, context drift, and logical inconsistency, improving scenario realism and compliance.	1 2 3 4
Hallucination & Context Drift Mitigation	RAG, multi-agent frameworks, and cognitive architectures reduce hallucinations and context drift, but complete elimination is infeasible; human-in-the-loop remains essential.	5 6 7 8
Evaluation Frameworks	New benchmarks (GraphEval, SECURE, CSEBenchmark, Drowzee) and metrics systematically measure hallucination rates and logical consistency in LLM-generated scenarios.	6 9 10
Security & Compliance	Blockchain, metadata management, adversarial tuning, and privacy controls are integrated to ensure compliance with DORA/TIBER-EU and system trustworthiness.	4 11 12 13
Research Gaps	Lack of unified architectures, domain-specific datasets, and scalable explainable models; further interdisciplinary research needed.	14 15 16 17

Direct Answer

Recent state-of-the-art research (2023-2025) on automating cyber range scenario generation using LLMs highlights the adoption of hybrid approaches—combining neuro-symbolic AI, Knowledge Graphs, and Multi-Agent Systems—to tackle critical challenges such as hallucinations, context drift, and maintaining logical consistency during threat simulation. These studies leverage advanced techniques such as Retrieval-Augmented Generation (RAG) to ground outputs in external, structured data, and incorporate multi-agent frameworks that facilitate cross-verification among different LLM outputs. Furthermore, a strong emphasis

is placed on compliance with cybersecurity frameworks like DORA and TIBER-EU, achieved by implementing robust security measures including data sanitization, provenance tracking, and blockchain-based audit trails. Despite these advancements, challenges remain in creating scalable, unified architectures and developing specialized datasets to benchmark the performance and reliability of these systems.

Study Scope

- **Time Period:** 2023–2025
- **Disciplines:** Cybersecurity, Artificial Intelligence, Machine Learning, Regulatory Compliance
- **Methods:** Meta-analysis of empirical studies, architectural reviews, benchmark evaluations, and synthesis of hybrid AI frameworks

Assumptions & Limitations

- **Assumptions:** The report assumes the continued evolution of LLMs and hybrid AI methods, and that regulatory frameworks (DORA, TIBER-EU) remain central to scenario generation requirements.
- **Limitations:** Limited by the availability of large-scale, domain-specific datasets and the nascent state of unified hybrid architectures; some findings are based on early-stage or pilot implementations.

Suggested Further Research

- Development of unified, scalable architectures integrating neuro-symbolic reasoning, multi-agent systems, and knowledge graphs.
- Creation of comprehensive, cybersecurity-specific benchmarks and datasets for hallucination and logical consistency evaluation.
- Exploration of blockchain and distributed ledger technologies for auditability and trust in automated scenario generation.
- Cross-industry collaboration to address explainability, security, and operational challenges in cyber range automation.

1. Introduction

Automated cyber range scenario generation is a cornerstone of modern cybersecurity training, enabling organizations to simulate complex threat environments and test resilience against evolving attacks. Large

Language Models (LLMs) have emerged as powerful tools for automating the creation of diverse and realistic threat scenarios, but their deployment introduces challenges such as hallucinations (fabricated or misleading outputs), context drift (loss of scenario coherence), and logical inconsistency. Addressing these issues is critical for effective training and regulatory compliance, especially under frameworks like DORA and TIBER-EU. Recent research converges on hybrid AI approaches—integrating neuro-symbolic reasoning, knowledge graphs, and multi-agent systems—to enhance scenario fidelity, explainability, and compliance.

Background and Motivation

The evolution from static, manually crafted cyber range scenarios to dynamic, LLM-driven automation reflects the need for scalable, adaptive, and realistic training environments. LLMs, particularly transformer-based models, offer generative capabilities that can simulate sophisticated attack chains and adversarial behaviors. However, their propensity for hallucinations and context drift poses risks in high-stakes cybersecurity applications. Hybrid AI methods, combining neural and symbolic reasoning, structured knowledge, and agentic workflows, are increasingly adopted to mitigate these challenges and ensure alignment with regulatory standards [5] [18] [19].

2. Advancements in LLM-Based Cyber Range Scenario Generation

Recent LLM Architectures and Domain-Specific Variants

- **Transformer-based Models:** The field has largely converged on decoder-only transformer architectures (e.g., LLaMA, GPT-4o), which balance scalability and efficiency for deployment in diverse environments, including resource-constrained edge devices [20].
- **Domain-Specific LLMs:** Models like GPT-4o and ICARuS demonstrate superior performance in automating security-critical tasks, leveraging curated datasets and advanced training methods to generate realistic cyber range scenarios [21] [22].
- **Efficiency Innovations:** Techniques such as Big-Little networks, implicit weight generation, and memory-centric architectures reduce computational overhead by up to 95%, facilitating practical deployment in organizations with limited resources [23] [24].

Retrieval-Augmented Generation (RAG) for Enhanced Scenario Realism

- **Dynamic Knowledge Integration:** RAG frameworks enable LLMs to retrieve and incorporate relevant external knowledge, improving factual consistency and reasoning accuracy for complex, multi-step scenarios [1] [25].
- **Graph-Based Retrieval:** Graph-based RAG techniques represent documents and relationships in graph structures, reducing hallucinations and supporting multi-hop reasoning for intricate scenario generation [26].

- **Self-Awareness Mechanisms:** Multi-round retrieval strategies (e.g., SIM-RAG) allow systems to determine when sufficient information has been gathered, enhancing the relevance and accuracy of generated scenarios [27].

Evaluation and Human-in-the-Loop Validation

- **Expert Review:** Iterative testing and human oversight are essential to ensure the realism, applicability, and alignment of LLM-generated scenarios with evolving threat landscapes [28] [29] [30].
- **Scenario Quality:** Evaluations reveal strong capabilities in automating malware detection and analysis, but variability persists in generating fully functional and diverse training content, underscoring the need for continued human-in-the-loop validation [31].

Synthesis

Recent advancements in LLM architectures and RAG frameworks have significantly improved the realism and adaptability of automated cyber range scenario generation. However, human oversight remains critical to address residual challenges in scenario quality and relevance.

3. Challenges in LLM Scenario Generation: Hallucinations, Context Drift, and Logical Consistency

Nature and Impact of Hallucinations and Context Drift

- **Hallucinations:** LLMs may generate fabricated, inaccurate, or misleading information, misidentifying benign activities as threats or missing real threats, leading to vulnerabilities and resource misallocation [5] .
- **Context Drift:** Loss of scenario coherence during multi-step generation arises from the vast and dynamic action space, inter-application cooperation, and constraint adherence, which current LLM agents struggle to maintain [32].
- **Logical Consistency:** Ensuring that generated scenarios adhere to logical constraints and domain-specific rules is a persistent challenge, especially in complex, evolving threat environments.

Mitigation Strategies and Hybrid Approaches

- **Hybrid Cognitive Architectures:** Integration of long-term memory, structured reasoning, and multi-modal learning reduces hallucinations and omissions compared to standalone LLMs [2].
- **Knowledge Graphs:** Provide structured factual context, filling knowledge gaps and enhancing reliability, though integration and evaluation methods remain active research areas [33] [34].
- **RAG and Advanced Prompting:** Retrieval-augmented generation, advanced prompting techniques, and

human oversight are effective in reducing hallucinations and improving logical consistency [29] [30] [35].

- **Multi-Agent Frameworks:** Adversarial debate and voting mechanisms enable cross-verification among LLMs, reducing hallucinations and improving accuracy and consistency [36].
- **Inherent Limitations:** Hallucinations are an unavoidable feature of LLMs due to fundamental computational and logical constraints; complete elimination is impossible, emphasizing mitigation and oversight [37].

Quantitative Evaluation Frameworks

- **Benchmarks:** GraphEval, SECURE, CSEBenchmark, and Drowzee provide systematic evaluation of hallucination rates and logical consistency in LLM-generated scenarios [6] [9] [10].
- **Metrics:** Mechanistic frameworks categorize hallucinations (factuality errors, faithfulness violations, logical inconsistencies), enabling precise identification and mitigation [38].
- **Dialogue-Level Evaluation:** DiaHalu and HalluScope benchmarks assess multi-turn hallucination and faithfulness, supporting fine-grained analysis of LLM outputs [39] [40].

Synthesis

While hybrid approaches and advanced evaluation frameworks have reduced hallucinations and context drift, these challenges remain inherent to LLMs. Systematic benchmarking and human oversight are essential for maintaining scenario quality and logical consistency.

4. Hybrid Neuro-Symbolic, Knowledge Graph, and Multi-Agent Approaches

Neuro-Symbolic AI Frameworks for Cybersecurity

- **Integration of Neural and Symbolic Reasoning:** Neuro-symbolic frameworks combine deep learning with symbolic reasoning, enhancing explainability, safety, and accuracy in cybersecurity applications [3] [7].
- **Graph Neural Networks (GNNs):** Embedding symbolic representations as graphs processed by GNNs enables multi-layer reasoning, capturing complex relational dependencies and improving expressivity [41].
- **Trade-offs:** Multistage symbolic integration improves explainability and robustness but involves a trade-off between formal logic precision and representational flexibility [17].

Knowledge Graph Integration and Explainability

- **Structured Context:** Knowledge graphs provide structured, domain-specific context, reducing

hallucinations and supporting interpretable reasoning in LLM-based systems [42] [43].

- **Semantic Enrichment:** Integration of ontologies and semantic enrichment in retrieval modules improves relevance and security of information accessed by LLMs [3].

Multi-Agent Systems and Reinforcement Learning

- **Agentic Workflows:** Multi-agent reinforcement learning systems combine neural and symbolic policies, enabling robust, interpretable, and adaptive threat simulation [44] [45].
- **Hierarchical Architectures:** Decompose cyber defense strategies into specialized subtasks coordinated by master policies, facilitating efficient adaptation to dynamic attacker strategies [45].
- **Ethical Constraints:** Integration of symbolic moral judging agents ensures adherence to ethical and regulatory constraints in automated scenario generation [46].

Trade-offs in Multistage Symbolic Integration

- **Precision vs. Flexibility:** Formal logic offers higher precision and verifiability, while representational-symbolic languages provide greater interpretability and adaptability to complex, real-world data [17] [47].
- **Scalability Challenges:** Exact formal verification is computationally intractable, necessitating approximate methods that balance guarantees with scalability [48] [49].

Synthesis

Hybrid neuro-symbolic, knowledge graph, and multi-agent approaches have substantially improved the accuracy, reliability, and explainability of automated cyber range scenario generation. However, balancing formal precision with representational flexibility and scalability remains an ongoing challenge.

5. Ensuring Compliance with Cybersecurity Frameworks: DORA and TIBER-EU

Automated Compliance Assessment and Scenario Generation

- **LLM-Driven Automation:** LLMs automate compliance assessment by analyzing unstructured organizational data, providing comprehensive evaluations of cybersecurity controls and supporting regulatory adherence [50] [51].
- **Scenario Quality:** Automated generation of realistic attack simulations and synthetic data validates security controls and response capabilities in line with DORA and TIBER-EU [52].

Hybrid LLM-RAG Systems for Continuous Compliance

- **Robust Security Measures:** Data filtering, sanitization, provenance tracking, adversarial training, and output monitoring prevent data poisoning, model manipulation, and privacy leakage, maintaining system integrity and reliability [4].
- **Verification Frameworks:** Continuum-Based Failure Classification (CBFC) models enable continuous, graded assessment of outputs for correctness, consistency, and uncertainty [53].
- **Multi-Agent Compliance:** Structured conversations and debate rounds operationalize ethical and legal compliance, aligning generated content with regulatory guidelines [54].

Security and Privacy Mitigation Strategies

- **Metadata Management:** Robust metadata management and data augmentation address cold-start challenges and improve system resilience against poisoning attacks [11].
- **Traceback Systems:** RAGForensics iteratively identifies poisoned texts, providing practical defense against sophisticated attacks [55].
- **Federated Learning:** Detection mechanisms measuring gradient differences and malicious client reporting defend against data poisoning and model manipulation [56].
- **Human-in-the-Loop:** Black-box defense frameworks combining zero-shot classifiers with human oversight enhance security against jailbreaking and manipulation [12].

Synthesis

Automated scenario generation techniques using LLMs and hybrid AI approaches have advanced compliance with DORA and TIBER-EU by integrating robust security, verification, and privacy controls. Continuous evaluation and human oversight remain essential for maintaining regulatory alignment.

6. Research Gaps and Emerging Trends

Current Research Gaps

- **Unified Architectures:** Lack of unified, dynamic architectures that seamlessly integrate factual grounding, scalability, and cybersecurity-specific training datasets [14] [15].
- **Benchmarking:** Scarcity of large-scale, domain-specific benchmarks and cross-domain evaluation frameworks for hallucination and logical consistency [14].
- **Explainability:** Challenges in scaling hybrid models with robust explainability and operational security, especially for non-expert users [17].

Emerging Hybrid AI Methods

- **Explainable AI (XAI):** Integration of XAI techniques (SHAP, LIME, Grad-CAM) into hybrid models improves interpretability, trust, and regulatory compliance [57] [58] [59].
- **Blockchain Integration:** Blockchain-based distributed ledger technologies provide tamper-resistant audit trails, enhancing trustworthiness and auditability in automated scenario generation [13] [60] [61].
- **Human-in-the-Loop Systems:** Interactive, human-AI workflows enable iterative refinement and vendor-agnostic deployment, addressing complexity and resource intensity [16].

Recent Datasets and Benchmarks

- **SECURE Benchmark:** Evaluates LLMs on realistic cybersecurity scenarios, focusing on knowledge extraction, understanding, and reasoning [10].
- **CSEBenchmark:** Fine-grained evaluation framework based on expert knowledge points and tailored questions [62].
- **Drowzee:** Logic-programming-aided metamorphic testing for fact-conflicting hallucination detection [6].
- **CyGPT:** Integrates cybersecurity-specific knowledge graphs with LLMs to enhance domain-specific reasoning and reduce hallucinations [63].

Explainable AI (XAI) Integration

- **Model-Agnostic Explanations:** Combining multiple XAI techniques provides comprehensive insights and improves interpretability in AI models [64].
- **Attention Mechanisms:** Tailored to complex AI-driven systems to enhance interpretability, regulatory compliance, and user trust [65] [66].

Blockchain for Trust and Auditability

- **Immutable Records:** Blockchain provides transparent, tamper-proof logging of cyber events and scenario executions, supporting effective incident detection and forensic analysis [13] [67].
- **Digital Twin Integration:** Enhances synchronization between physical and virtual assets, ensuring secure data sharing and lifecycle security [68] [69].

Synthesis

Emerging trends emphasize hybrid AI methods, explainable frameworks, and blockchain integration to address trustworthiness, explainability, and auditability gaps in automated cyber range scenario generation.

Continued interdisciplinary research and cross-industry collaboration are needed to advance these solutions.

7. Conclusion

Summary of Key Insights

- **Hybrid AI Approaches:** Integration of neuro-symbolic reasoning, knowledge graphs, and multi-agent systems with LLMs has improved the fidelity, security, and compliance of automated cyber range scenario generation [3] [5] [18] [28] [70].
- **Mitigation of Hallucinations and Context Drift:** RAG, cognitive architectures, and multi-agent frameworks have reduced hallucinations and context drift, but complete elimination remains infeasible; human oversight is essential.
- **Compliance and Security:** Robust security measures, verification frameworks, and blockchain-based audit trails ensure alignment with DORA and TIBER-EU, enhancing trustworthiness and auditability.
- **Research Gaps:** Unified architectures, specialized benchmarks, and scalable explainable models are needed to further advance the field.

Future Directions

- **Unified Architectures:** Develop scalable, unified frameworks integrating neuro-symbolic reasoning, multi-agent systems, and knowledge graphs for robust scenario generation [14] [16] [71].
- **Benchmarking and Evaluation:** Create comprehensive, cybersecurity-specific datasets and evaluation frameworks for hallucination and logical consistency.
- **Blockchain and Distributed Ledger Technologies:** Explore blockchain integration for enhanced auditability and trust in automated systems.
- **Interdisciplinary Collaboration:** Foster cross-industry partnerships to address explainability, security, and operational challenges in cyber range automation.

Synthesis

Recent advancements in automating cyber range scenario generation using LLMs are marked by a shift towards hybrid AI methods that integrate neuro-symbolic reasoning, knowledge graphs, and multi-agent systems. These approaches have improved the fidelity, security, and compliance of simulated cyber threat scenarios, particularly addressing persistent challenges such as hallucinations, context drift, and logical inconsistencies. Despite these innovations, there remain critical research gaps in scaling these solutions and developing domain-specific benchmarks, signifying a need for further interdisciplinary studies that combine rigorous evaluation metrics with robust, secure architectures.

References

1. CRP-RAG: A Retrieval-Augmented Generation Framework for Supporting Complex Logical Reasoning and Knowledge Planning Kehan, X., Kun, Z., Jingyuan, X., Wei, H. Electronics (Switzerland), 2025 https://www.scopus.com/pages/publications/85214446183?origin=scopusAI

2. Filtering Hallucinations and Omissions in Large Language Models through a Cognitive Architecture Asaduzzaman, M., Giorgi, I., Masala, G.L. 2025 IEEE Symposium on Computational Intelligence in Natural Language Processing and Social Media, CI-NLPSoMe Companion 2025, 2025 https://www.scopus.com/pages/publications/105005026731?origin=scopusAI

3. Knowledge-Enhanced Neurosymbolic Artificial Intelligence for Cybersecurity and Privacy Piplai, A., Kotal, A., Mohseni, S., (...), Joshi, A. IEEE Internet Computing, 2023 https://www.scopus.com/pages/publications/85174520555?origin=scopusAI

4. Enhancing Communication and Data Transmission Security in RAG Using Large Language Models Gummadi, V., Udayaraju, P., Sarabu, V.R., (...), Venkataramana, S. 4th International Conference on Sustainable Expert Systems, ICSES 2024 - Proceedings, 2024 https://www.scopus.com/pages/publications/85214782970?origin=scopusAI

5. The Paradigm of Hallucinations in AI-driven cybersecurity systems: Understanding taxonomy, classification outcomes, and mitigations Sood, A.K., Zeadally, S., Hong, E. Computers and Electrical Engineering, 2025 https://www.scopus.com/pages/publications/105001875892?origin=scopusAI

6. Drowzee: Metamorphic Testing for Fact-Conflicting Hallucination Detection in Large Language Models Li, N., Li, Y., Liu, Y., (...), Wang, H. Proceedings of the ACM on Programming Languages, 2024 https://www.scopus.com/pages/publications/85206945165?origin=scopusAI

7. NSCTI: A Hybrid Neuro-Symbolic Framework for AI-Driven Predictive Cyber Threat Intelligence Nalluri, S., Malyala, M.M., Kandagiri, H., Kandagiri, K.K. Proceedings - 2025 4th International Conference on Computational Modelling, Simulation and Optimization, ICCMSO 2025, 2025 https://www.scopus.com/pages/publications/105013839985?origin=scopusAI

8. ATM: Adversarial Tuning Multi-agent System Makes a Robust Retrieval-Augmented GENERATOR Zhu, J., Yan, L., Shi, H., (...), Sha, L. EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2024 https://www.scopus.com/pages/publications/85217818171?origin=scopusAI

9. GraphEval: A Knowledge-Graph Based LLM Hallucination Evaluation Framework Sansford, H., Richardson, N., Maretic, H.P., Saada, J.N. CEUR Workshop Proceedings, 2024 https://www.scopus.com/pages/publications/85216407923?origin=scopusAI

10. SECURE: Benchmarking Large Language Models for Cybersecurity Bhusal, D., Alam, M.T., Nguyen, L., (...), Rastogi, N. Proceedings - Annual Computer Security Applications Conference, ACSAC, 2024 https://www.scopus.com/pages/publications/105001234955?origin=scopusAI

11. Poison-RAG: Adversarial Data Poisoning Attacks on Retrieval-Augmented Generation in Recommender Systems Nazary, F., Deldjoo, Y., Noia, T.D. Lecture Notes in Computer Science, 2025 https://www.scopus.com/pages/publications/105006569504?origin=scopusAI

12. LLM-Sentry: A Model-Agnostic Human-in-the-Loop Framework for Securing Large Language Models Irtiza, S., Akbar, K.A., Yasmeen, A., (...), Thuraisingham, B. Proceedings - 2024 IEEE 6th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications, TPS-ISA 2024, 2024 https://www.scopus.com/pages/publications/85217836064?origin=scopusAI

13. Advanced Cyber Security Techniques for Data, Blockchain, IoT, and Network Protection Chaubey, N.K., Chaubey, N. Advanced Cyber Security Techniques for Data, Blockchain, IoT, and Network Protection, 2024 https://www.scopus.com/pages/publications/105014055148?origin=scopusAI

14. A Systematic Literature Review of Hallucinations in Large Language Models Woesle, C., Fischer-Brandies, L., Buettner, R. IEEE Access, 2025 https://www.scopus.com/pages/publications/105013999771?origin=scopusAI

15. Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities Ferrag, M.A., Alwahedi, F., Battah, A., (...), Debbah, M. Internet of Things and Cyber-Physical Systems, 2025 https://www.scopus.com/pages/publications/105004989357?origin=scopusAI

16. From Concept to Deployment: An AI Assistant for Generating and Configuring Cyber Range Scenarios Rizos, G.S., Kopalidis, N., Mengidis, N., (...), Votis, K. Proceedings of the 2025 IEEE International Conference on Cyber Security and Resilience, CSR 2025, 2025 https://www.scopus.com/pages/publications/105016166909?origin=scopusAI

17. A Roadmap Toward Neurosymbolic Approaches in AI Design Arachchige, P.J., Iancu, B., Lilius, J. IEEE Access, 2025 https://www.scopus.com/pages/publications/105018075222?origin=scopusAI

18. LLM Inference Serving: Survey of Recent Advances and Opportunities Li, B., Jiang, Y., Gadepally, V., Tiwari, D. 2024 IEEE High Performance Extreme Computing Conference, HPEC 2024, 2024 https://www.scopus.com/pages/publications/105002728757?origin=scopusAI

19. Transitioning from MLOps to LLMOps: Navigating the Unique Challenges of Large Language Models Pahune, S., Akhtar, Z. Information (Switzerland), 2025 https://www.scopus.com/pages/publications/85218460051?origin=scopusAI

20. Survey and Evaluation of Converging Architecture in LLMs Based on Footsteps of Operations Kim, S., Moon, J., Oh, J., (...), Yang, J.-S. IEEE Open Journal of the Computer Society, 2025 https://www.scopus.com/pages/publications/105010893716?origin=scopusAI

21. LLMs for Microservice Generation: Capabilities, Challenges, and Advancements Spista, R., Crispo, B., Giorgini, P., (...), Riccardi, G. Proceedings of the 2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2025, 2025 https://www.scopus.com/pages/publications/105014314566?origin=scopusAI

22. ICARuS: Intercode-CTF Auto-Randomization System Kerr, R., Taylor, A., Sultana, M., El-Rami, J.-P.S. Proceedings - 2025 IEEE Conference on Artificial Intelligence, CAI 2025, 2025 https://www.scopus.com/pages/publications/105011261167?origin=scopusAI

23. 20.5 C-Transformer: A 2.6-18.1μJ/Token Homogeneous DNN-Transformer/Spiking-Transformer Processor with Big-Little Network and Implicit Weight Generation for Large Language Models Kim, S., Kim, S., Jo, W., (...), Yoo, H.-J. Digest of Technical Papers - IEEE International Solid-State Circuits Conference, 2024 https://www.scopus.com/pages/publications/85188051669?origin=scopusAI

24. Special Session: Neuro-Symbolic Architecture Meets Large Language Models: A Memory-Centric

Perspective Ibrahim, M., Wan, Z., Li, H., (...), Raychowdhury, A. Proceedings - 2024 International Conference on Hardware/Software Codesign and System Synthesis, CODES+ISSS 2024, 2024 https://www.scopus.com/pages/publications/85212298689?origin=scopusAI

25. Enhancing Language Models with Retrieval-Augmented Generation A Comparative Study on Performance Grabuloski, M., Karadimce, A., Sefidanoski, A. WSEAS Transactions on Information Science and Applications, 2025 https://www.scopus.com/pages/publications/105002172474?origin=scopusAI

26. Enhancing Document Retrieval Using AI and Graph-Based RAG Techniques Kamra, V., Gupta, L., Arora, D., Yadav, A.K. Proceedings - IEEE 5th International Conference on Communication, Computing and Industry 6.0 2024, C2I6 2024, 2024 https://www.scopus.com/pages/publications/105000431780?origin=scopusAI

27. Knowing You Don't Know: Learning When to Continue Search in Multi-round RAG through Self-Practicing Yang, D., Zeng, L., Rao, J., Zhang, Y. SIGIR 2025 - Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2025 https://www.scopus.com/pages/publications/105011818246?origin=scopusAI

28. Applications of LLMs for Generating Cyber Security Exercise Scenarios Mudassar Yamin, M., Hashmi, E., Ullah, M., Katt, B. IEEE Access, 2024 https://www.scopus.com/pages/publications/85205455626?origin=scopusAI

29. Strategies to mitigate hallucinations in large language models Bhattacharya, R. Applied Marketing Analytics, 2024 https://www.scopus.com/pages/publications/85202886920?origin=scopusAI

30. Understanding and Mitigating Hallucinations in Large Language Models: Insights from a Systematic Literature Review Singh, R., Singh, P., Malik, A., Sukmawan, D. 2025 International Conference on Metaverse and Current Trends in Computing, ICMCTC 2025, 2025 https://www.scopus.com/pages/publications/105022249897?origin=scopusAI

31. Evaluation of the maturity of LLMs in the cybersecurity domain Conceição, T., Cruz, N. International Journal of Information Security, 2025 https://www.scopus.com/pages/publications/105014594886?origin=scopusAI

32. Understanding the Weakness of Large Language Model Agents within a Complex Android Environment Xing, M., Zhang, R., Xue, H., (...), Xiao, Z. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2024 https://www.scopus.com/pages/publications/85203687387?origin=scopusAI

33. Knowledge Graphs, Large Language Models, and Hallucinations: An NLP Perspective Lavrinovics, E., Biswas, R., Bjerva, J., Hose, K. Journal of Web Semantics, 2025 https://www.scopus.com/pages/publications/85213832166?origin=scopusAI

34. DIVKNOWQA: Assessing the Reasoning Ability of LLMs via Open-Domain Question Answering over Knowledge Base and Text Zhao, W., Liu, Y., Niu, T., (...), Yavuz, S. Findings of the Association for Computational Linguistics: NAACL 2024 - Findings, 2024 https://www.scopus.com/pages/publications/85197851513?origin=scopusAI

35. Mitigating Hallucinations in Large Language Models: A Comprehensive Survey on Detection and Reduction Strategies Saxena, V., Sathe, A., Sandosh, S. Lecture Notes in Networks and Systems, 2025 https://www.scopus.com/pages/publications/105005480229?origin=scopusAI

36. Minimizing Hallucinations and Communication Costs: Adversarial Debate and Voting Mechanisms in LLM-Based Multi-Agents Yang, Y., Ma, Y., Feng, H., (...), Han, Z. Applied Sciences (Switzerland), 2025 https://www.scopus.com/pages/publications/105002274951?origin=scopusAI

37. LLMs Will Always Hallucinate, and We Need to Live with This Banerjee, S., Agarwal, A., Singla, S. Lecture Notes in Networks and Systems, 2025 https://www.scopus.com/pages/publications/105017227681?origin=scopusAI

38. Hallucination in Large Language Models: From Mechanistic Understanding to Novel Control Frameworks Zhang, W., Zhang, C., Gu, C., (...), Fang, Y. 7th International Conference on Universal Village, UV 2024, 2024 https://www.scopus.com/pages/publications/105021208926?origin=scopusAI

39. DiaHalu: A Dialogue-level Hallucination Evaluation Benchmark for Large Language Models Chen, K., Chen, Q., Zhou, J., (...), He, L. EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing, Findings of EMNLP 2024, 2024 https://www.scopus.com/pages/publications/85217618244?origin=scopusAI

40. HalluScope: A Comprehensive Dataset for Evaluating Hallucination in Large Language Models Across Multiple Domains Zhao, C., Zeng, B., Chen, K., Lin, X. Communications in Computer and Information Science, 2025 https://www.scopus.com/pages/publications/105012430042?origin=scopusAI

41. DEEPGRAPHLOG for Layered Neurosymbolic AI Kikaj, A., Marra, G., Geerts, F., (...), De Raedt, L. Frontiers in Artificial Intelligence and Applications, 2025 https://www.scopus.com/pages/publications/105024480603?origin=scopusAI

42. Integrating Knowledge Graphs with Symbolic AI: The Path to Interpretable Hybrid AI Systems in Medicine Vidal, M.-E., Chudasama, Y., Huang, H., (...), Torrente, M. Journal of Web Semantics, 2025 https://www.scopus.com/pages/publications/85213232059?origin=scopusAI

43. Formal Knowledge Augmented Language Models for Explainable and Robust Reasoning Nguyen, N.-K., Nguyen, V.-H., Le, A.-C. Journal of Intelligent and Fuzzy Systems, 2025 https://www.scopus.com/pages/publications/105024216138?origin=scopusAI

44. BLENDRL: A FRAMEWORK FOR MERGING SYMBOLIC AND NEURAL POLICY LEARNING Shindo, H., Delfosse, Q., Dhami, D.S., Kersting, K. 13th International Conference on Learning Representations, ICLR 2025, 2025 https://www.scopus.com/pages/publications/105010233028?origin=scopusAI

45. Hierarchical Multi-agent Reinforcement Learning for Cyber Network Defense: Extended Abstract Singh, A.V., Rathbun, E., Graham, E., (...), Oprea, A. Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, 2025 https://www.scopus.com/pages/publications/105009810417?origin=scopusAI

46. AJAR: An Argumentation-based Judging Agents Framework for Ethical Reinforcement Learning Alcaraz, B., Chaput, R., Boissier, O., Leturc, C. Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, 2023 https://www.scopus.com/pages/publications/85171276678?origin=scopusAI

47. Neuro-Symbolic AI: A Future of Tomorrow Chandre, P., Mahalle, P., Shinde, G., (...), Kashid, S. ASEAN Journal on Science and Technology for Development, 2025 https://www.scopus.com/pages/publications/105003053014?origin=scopusAI

48. Efficient Neuro-Symbolic Policy using In-Memory Computing Molom-Ochir, T., Saxena, N., Kim, J., (...), Helen, H. Proceedings of Machine Learning Research, 2025 https://www.scopus.com/pages/publications/105014733669?origin=scopusAI

49. A Scalable Approach to Probabilistic Neuro-Symbolic Robustness Verification Manginas, V., Manginas, N., Stevinson, E., (...), Lomuscio, A. Proceedings of Machine Learning Research, 2025 https://www.scopus.com/pages/publications/105020242224?origin=scopusAI

50. Position Paper: Leveraging Large Language Models for Cybersecurity Compliance Salman, A., Creese, S., Goldsmith, M. Proceedings - 9th IEEE European Symposium on Security and Privacy Workshops, Euro S and PW 2024, 2024 https://www.scopus.com/pages/publications/85203022174?origin=scopusAI

51. Exploring Language Agents for Automated Compliance Audits in Cybersecurity Governance Huang, A., Chao, A. ICCE-Taiwan 2025 - 12th IEEE International Conference on Consumer Electronics - Taiwan: Generative AI in Innovative Consumer Technology, Proceedings, 2025 https://www.scopus.com/pages/publications/105022419221?origin=scopusAI

52. Security Copilot: A Blueprint for Leveraging Generative AI in Cyber Defense Krishnan, V.V., Sudha, V.K. Lecture Notes in Networks and Systems, 2025 https://www.scopus.com/pages/publications/105006917273?origin=scopusAI

53. Verification and Validation of LLM-RAG for Industrial Automation Min, Z., Budnik, C.J. Proceedings - 2025 IEEE International Conference on Artificial Intelligence Testing, AITest 2025, 2025 https://www.scopus.com/pages/publications/105016139401?origin=scopusAI

54. Grounded Ethical AI: A Demonstrative Approach with RAG-Enhanced Agents de Cerqueira, J.A.S., Khan, A.A., Rousi, R., (...), Abrahamsson, P. CEUR Workshop Proceedings, 2025 https://www.scopus.com/pages/publications/85218445480?origin=scopusAI

55. Traceback of Poisoning Attacks to Retrieval-Augmented Generation Zhang, B., Xin, H., Fang, M., (...), Liu, Z. WWW 2025 - Proceedings of the ACM Web Conference, 2025 https://www.scopus.com/pages/publications/105005159200?origin=scopusAI

56. DM-FedMF: A Recommendation Model of Federated Matrix Factorization With Detection Mechanism Zheng, X., Jia, X., Cheng, X., (...), Luo, Y. IEEE Transactions on Network Science and Engineering, 2025 https://www.scopus.com/pages/publications/105000657710?origin=scopusAI

57. A Hybrid Explainable AI Framework for Enhancing Trust and Transparency in Autonomous Vehicles Shinde, R.K., Shinde, K.D., Mehta, H. 2025 International Conference on Emerging Smart Computing and Informatics, ESCI 2025, 2025 https://www.scopus.com/pages/publications/105007282568?origin=scopusAI

58. Explainable AI in Cybersecurity: A Comprehensive Framework for enhancing transparency, trust, and Human-AI Collaboration Desai, B., Patil, K., Mehta, I., Patil, A. Proceedings - 2024 International of Seminar on Application for Technology of Information and Communication: Smart And Emerging Technology for a Better Life, iSemantic 2024, 2024 https://www.scopus.com/pages/publications/85213368431?origin=scopusAI

59. Towards White-Box IDS: Integrating Explainability in LoT Ecosystems Tewari, T., Singal, G. Proceedings of the National Conference on Communications, NCC, 2025 https://www.scopus.com/pages/publications/105010681741?origin=scopusAI

60. ZAIA: Zero-trust Authentication and Identity Attestation Framework for AI-Enabled IIoTs in Smart

Manufacturing Ecosystem Verma, R., Indra, G. 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies, INSPECT 2024, 2024 https://www.scopus.com/pages/publications/105000691462?origin=scopusAI

61. CyberDetect MLP a big data enabled optimized deep learning framework for scalable cyberattack detection in IoT environments Upender, T., Neelakantappa, M., Rao, C.P., (...), Yamsani, N. Scientific Reports, 2025 https://www.scopus.com/pages/publications/105022223896?origin=scopusAI

62. The Digital Cybersecurity Expert: How Far Have We Come? Wang, D., Zhou, G., Li, X., (...), Li, D. Proceedings - IEEE Symposium on Security and Privacy, 2025 https://www.scopus.com/pages/publications/105009322568?origin=scopusAI

63. CyGPT: Knowledge Graph-Based Enhancement Techniques for Large Language Models in Cybersecurity Ou, L., Ni, X., Wu, W., Tian, Z. Proceeding - 2024 IEEE 9th International Conference on Data Science in Cyberspace, DSC 2024, 2024 https://www.scopus.com/pages/publications/85218447683?origin=scopusAI

64. Understanding explainable artificial intelligence techniques: a comparative analysis for practical application Bhatnagar, S., Agrawal, R. Bulletin of Electrical Engineering and Informatics, 2024 https://www.scopus.com/pages/publications/85205273476?origin=scopusAI

65. Explainable Artificial Intelligence (XAI) for Trustworthy AI in 6G Networks Shafik, W. 6G Networks and AI-Driven Cybersecurity, 2025 https://www.scopus.com/pages/publications/105015932749?origin=scopusAI

66. Explainable AI (XAI): Techniques, applications, and challenges Kopzhasarova, M., Kozhamzharova, D. CEUR Workshop Proceedings, 2025 https://www.scopus.com/pages/publications/105006886912?origin=scopusAI

67. Digital Twin-Enabled Incident Detection and Response: A Systematic Review of Critical Infrastructures Applications Kampourakis, K.E., Gkioulos, V., Kavallieratos, G., Lin, J.-C. International Journal of Information Security, 2025 https://www.scopus.com/pages/publications/105013958140?origin=scopusAI

68. Reserach on Digital Twins Technology in Cyberspace Security Qiankun, R., Xinli, X., Jingju, L., Qian, Y. Xitong Fangzhen Xuebao / Journal of System Simulation, 2024 https://www.scopus.com/pages/publications/85202055322?origin=scopusAI

69. Digital Twin in Industries: A Comprehensive Survey Bokhtiar Al Zami, M., Shaon, S., Khanh Quy, V., Nguyen, D.C. IEEE Access, 2025 https://www.scopus.com/pages/publications/105001207066?origin=scopusAI

70. Emerging trends in hybrid information systems modeling in artificial intelligence Sakshi, Mehrotra, T., Tyagi, P., Jain, V. Hybrid Information Systems: Non-Linear Optimization Strategies with Artificial Intelligence, 2024 https://www.scopus.com/pages/publications/85199937050?origin=scopusAI

71. AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges Córdova-Esparza, D.-M. Information (Switzerland), 2025 https://www.scopus.com/pages/publications/105009051718?origin=scopusAI