

DAFTAR PUSTAKA

- Abdurrasyid, Asnar, Y., & Ayu Putri Saptawati, G. (2026). Mutation based improvement of security test case quality for broken access control. *Journal on Information Security*, 2026(1), 7. <https://doi.org/10.1186/s13635-026-00226-w>
- Akhoundali, J., Hamidi, H., Rietveld, K., & Gadyatskaya, O. (2025). Eradicating the Unseen: Detecting, Exploiting, and Remediating a Path Traversal Vulnerability across GitHub. *Proceedings of the ACM Conference on Computer and Communications Security*, 542–558. <https://doi.org/10.1145/3708821.3736220>
- Alhuzali, A. (2025). LLM-powered threat intelligence: A retrieval-augmented generation approach for cyber attack investigation. *PeerJ Computer Science*, 11. <https://doi.org/10.7717/peerj-cs.3371>
- Altulaihan, E. A., Alismail, A., & Frikha, M. (2023). A Survey on Web Application Penetration Testing. In *Electronics (Switzerland)* (Vol. 12, Number 5). MDPI. <https://doi.org/10.3390/electronics12051229>
- Anas, A., Alhelbawy, A. A., Gamal, S. El, & Youssef, B. (2024). BACAD: AI-based framework for detecting vertical broken access control attacks. *Egyptian Informatics Journal*, 28. <https://doi.org/10.1016/j.eij.2024.100571>
- Antal, M., & Buza, K. (2025). Evaluating Open-Source LLMs in RAG Systems: A Benchmark on Diploma Theses Abstracts Using Ragas. *Acta Universitatis Sapientiae, Informatica*, 17(1). <https://doi.org/10.1007/s44427-025-00006-3>
- Arslan, M., Ghanem, H., Munawar, S., & Cruz, C. (2024). A Survey on RAG with LLMs. *Procedia Computer Science*, 246(C), 3781–3790. <https://doi.org/10.1016/j.procs.2024.09.178>
- Asri, Y., Kuswardani, D., Nita Suliyanti, W., & Monica Tambunan, C. (2022). Classification of Points and Types of Disruptions for Efficiency of Standard Operating Procedures Repair of Distribution Network Suppliers. *AIP Conference Proceedings*, 2499. <https://doi.org/10.1063/5.0104937>
- Baysan, M. S., Uysal, S., İşlek, İ., Çiğ Karaman, Ç., & Güngör, T. (2025). LLM-as-a-Judge: automated evaluation of search query parsing using large language models. *Frontiers in Big Data*, 8. <https://doi.org/10.3389/fdata.2025.1611389>
- Blefari, F., Cosentino, C., Pironti, F. A., Furfaro, A., & Marozzo, F. (2026). CyberRAG: An agentic RAG cyber attack classification and reporting tool. *Future Generation Computer Systems*, 176. <https://doi.org/10.1016/j.future.2025.108186>
- Bora, A., & Cuayáhuítl, H. (2024). Systematic Analysis of Retrieval-Augmented Generation-Based LLMs for Medical Chatbot Applications. *Machine Learning and Knowledge Extraction*, 6(4), 2355–2374. <https://doi.org/10.3390/make6040116>

- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language Models are Few-Shot Learners*.
<http://arxiv.org/abs/2005.14165>
- Desyana, N., & Luthfiarta, A. (2025). Analisis Komparatif Kinerja Llama Murni, Rag Native, dan Rag Fine-Tuning. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 11(3), 263–272. <https://doi.org/10.25077/teknosi.v11i3.2025.263-272>
- Du, X., Zheng, G., Wang, K., Zou, Y., Wang, Y., Deng, W., Feng, J., Liu, M., Chen, B., Peng, X., Ma, T., & Lou, Y. (2026). Vul-RAG: Enhancing LLM-based Vulnerability Detection via Knowledge-level RAG. *ACM Transactions on Software Engineering and Methodology*. <https://doi.org/10.1145/3797277>
- Durmuşkaya, M. E., & Bayraklı, S. (2025). Web application firewall based on machine learning models. *PeerJ Computer Science*, 11. <https://doi.org/10.7717/peerj-cs.2975>
- Elzemity, A., Arief, B., & Li, S. (2025). CyberLLMInstruct: A Pseudo-Malicious Dataset Revealing Safety-Performance Trade-offs in Cyber Security LLM Fine-tuning. *Proceedings of the 18th ACM Workshop on Artificial Intelligence and Security, AISEC 2025*, 77–88. <https://doi.org/10.1145/3733799.3762968>
- Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2024). RAGAS: Automated Evaluation of Retrieval Augmented Generation. <https://doi.org/https://doi.org/10.18653/v1/2024.eacl-demo.16>
- Farzaan, M. A. M., Ghanem, M. C., El-Hajjar, A., & Ratnayake, D. N. (2025). AI-Powered System for an Efficient and Effective Cyber Incidents Detection and Response in Cloud Environments. *IEEE Transactions on Machine Learning in Communications and Networking*, 3, 623–643.
<https://doi.org/10.1109/tmlcn.2025.3564912>
- Gomez-Cabello, C. A., Prabha, S., Haider, S. A., Genovese, A., Collaco, B. G., Wood, N. G., Bagaria, S., & Forte, A. J. (2025). Comparative Evaluation of Advanced Chunking for Retrieval-Augmented Generation in Large Language Models for Clinical Decision Support. *Bioengineering*, 12(11).
<https://doi.org/10.3390/bioengineering12111194>
- Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., Liu, H., Wang, S., Zhang, K., Lin, Z., Zhang, B., Ni, L., Gao, W., Wang, Y., & Guo, J. (2026). A survey on LLM-as-a-Judge. *The Innovation*, 101253.
<https://doi.org/10.1016/j.xinn.2025.101253>
- Guduru, S. (2025). Autonomous Cyber Defense: LLM-Powered Incident Response with LangChain and SOAR Integration. *International Journal of Computer Science and*

- Information Technology Research*, 6(1), 72–82.
https://doi.org/10.63530/ijcsitr_2025_06_01_008
- Hasanov, I., Virtanen, S., Hakkala, A., & Isoaho, J. (2024). Application of Large Language Models in Cybersecurity: A Systematic Literature Review. *IEEE Access*, 12, 176751–176778. <https://doi.org/10.1109/ACCESS.2024.3505983>
- Indrianto, Abdurrasyid, Susanti, M. N. I., & Ramadhan, A. (2023). Text-to-speech on health monitoring bracelet for the visually impaired. *Bulletin of Electrical Engineering and Informatics*, 12(6), 3826–3836.
<https://doi.org/10.11591/eei.v12i6.5369>
- Ismail, Kurnia, R., Brata, Z. A., Nelistiani, G. A., Heo, S., Kim, H., & Kim, H. (2025). Toward Robust Security Orchestration and Automated Response in Security Operations Centers with a Hyper-Automation Approach Using Agentic Artificial Intelligence. *Information (Switzerland)*, 16(5).
<https://doi.org/10.3390/info16050365>
- Ismail, Kurnia, R., Widyatama, F., Wibawa, I. M., Brata, Z. A., Ukasyah, Nelistiani, G. A., & Kim, H. (2025). Enhancing Security Operations Center: Wazuh Security Event Response with Retrieval-Augmented-Generation-Driven Copilot. *Sensors*, 25(3). <https://doi.org/10.3390/s25030870>
- Jaffal, N. O., Alkhanafseh, M., & Mohaisen, D. (2025). Large Language Models in Cybersecurity: A Survey of Applications, Vulnerabilities, and Defense Techniques. In *AI (Switzerland)* (Vol. 6, Number 9). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/ai6090216>
- Jones, G., Kasimatis, D., Pitropakis, N., Macfarlane, R., & Buchanan, W. J. (2025). Analysing the role of LLMs in cybersecurity incident management. *International Journal of Information Security*, 24(6). <https://doi.org/10.1007/s10207-025-01144-7>
- Klesel, M., & Wittmann, H. F. (2025). Retrieval-Augmented Generation (RAG). *Business and Information Systems Engineering*, 67(4), 551–561.
<https://doi.org/10.1007/s12599-025-00945-3>
- Kumar, A., & Gutierrez, J. A. (2025). Impact of Machine Learning on Intrusion Detection Systems for the Protection of Critical Infrastructure. In *Information (Switzerland)* (Vol. 16, Number 7). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/info16070515>
- Kusnandar, A. R., & Herman Bedi Agtriadi. (2025). Utilization of LLM in the Automation Process of Contract Template Recognition. *Jurnal E-Komtek (Elektro-Komputer-Teknik)*, 9(2), 563–581. <https://doi.org/10.37339/e-komtek.v9i2.2511>

- Lestari, M., Djunaidi, K., Rahma, ;, Ningrum, F., Dine, ;, Kusuma, T., Khesya, ;, & Saradiba, M. (2023). Implementasi Cosine Similarity dalam Pencarian Fungsional Obat. In *Prosiding Seminar Nasional Energi, Kelistrikan, Teknik dan Informatika* (Vol. 94).
- Lin, X., Zhang, J., Deng, G., Liu, T., Zhang, T., Guo, Q., & Chen, R. (2025). *IRCopilot: Automated Incident Response with Large Language Models*. <http://arxiv.org/abs/2505.20945>
- Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (n.d.). *Lost in the Middle: How Language Models Use Long Contexts*. <https://doi.org/10.1162/tacl>
- Liu, X., Wang, J., Yuan, X., Sun, J., Dong, G., Di, P., Wang, W., & Wang, D. (2026). Prompting Frameworks for Large Language Models: A Survey. *ACM Computing Surveys*, 58(10), 1–38. <https://doi.org/10.1145/3789253>
- Liu, Y. Y., Zheng, Z., Zhang, F., Feng, J. C., Fu, Y. Y., Zhai, J. D., He, B. S., Zhang, X., & Du, X. Y. (2026). A comprehensive taxonomy of prompt engineering techniques for large language models. In *Frontiers of Computer Science* (Vol. 20, Number 3). Higher Education Press Limited Company. <https://doi.org/10.1007/s11704-025-50058-z>
- Loumachi, F. Y., Ghanem, M. C., & Ferrag, M. A. (2025). Advancing Cyber Incident Timeline Analysis Through Retrieval-Augmented Generation and Large Language Models. *Computers*, 14(2). <https://doi.org/10.3390/computers14020067>
- Lucz, G., & Forstner, B. (2025). *A Thirty-Day Dataset of Malicious HTTP Requests Blocked by OWASP ModSecurity on a Production Web Server*. <https://doi.org/10.3390/data>
- Maniraj, S. P., Sabapathy Ranganathan, C., & Sekar, S. (2024). SECURING WEB APPLICATIONS WITH OWASP ZAP FOR COMPREHENSIVE SECURITY TESTING. In *Int. J. Adv. Sig. Img. Sci* (Vol. 10, Number 2).
- Microsoft. (n.d.). *Windows Subsystem for Linux Documentation*.
- MITRE. (2024). *CWE-22 : Path Traversal & CWE-23 : Reverse-Path Traversal*. MITRE.
- Nawalny, M., Łepicki, M., Latkowski, T., Bujak, S., Bukowski, M., Świdorski, B., Baranik, G., Nowak, B., Zakowicz, R., Dobrakowski, Ł., Oczeretko, A., Sadowski, P., Szlaga, K., Kubica, B., & Kurek, J. (2026). Comparative Evaluation of GPT-4o, GPT-OSS-120B and Llama-3.1-8B-Instruct Language Models in a Reproducible CV-to-JSON Extraction Pipeline. *Applied Sciences (Switzerland)*, 16(1). <https://doi.org/10.3390/app16010217>

- Nawrocki, M., & Kołodziej, J. (2024). Vulnerabilities of Web Applications: Good Practices and New Trends. *Applied Cybersecurity and Internet Governance*, 3(2), 122–143. <https://doi.org/10.60097/ACIG/199521>
- Nelson, A., Rekhi, S., Souppaya, M., & Scarfone, K. (2025). *Incident response recommendations and considerations for cybersecurity risk management* : <https://doi.org/10.6028/NIST.SP.800-61r3>
- Ouchebara, D. S., & Dupont, S. (2025). *Llama-based source code vulnerability detection: Prompt engineering vs Fine tuning*. <http://arxiv.org/abs/2512.09006>
- OWASP Contributors. (n.d.). *A01:2025 Broken Access Control*.
- Pasini, S., Kim, J., Aiello, T., Cabrera Lozoya, R., Sabetta, A., & Tonella, P. (2026). Evaluating and improving the robustness of security attack detectors generated by LLMs. *Empirical Software Engineering*, 31(2). <https://doi.org/10.1007/s10664-025-10743-w>
- Pingua, B., Sahoo, A., Kandpal, M., Murmu, D., Rautaray, J., Barik, R. K., & Saikia, M. J. (2025). Medical LLMs: Fine-Tuning vs. Retrieval-Augmented Generation. *Bioengineering*, 12(7). <https://doi.org/10.3390/bioengineering12070687>
- Radanliev, P. (2025). Collaborative penetration testing suite for emerging generative AI algorithms. *Applied Intelligence*, 55(16). <https://doi.org/10.1007/s10489-025-06908-1>
- Rajapaksha, S., Rani, R., & Karafili, E. (2024). *A RAG-Based Question-Answering Solution for Cyber-Attack Investigation and Attribution*. <http://arxiv.org/abs/2408.06272>
- Reyes Narváez, A., Curipallo Martínez, M., Reyes Narváez, E., Lara, F., Reyes Narváez, E. P., & Barba Molina, H. (2025). Evaluation Framework for False Positives in Open-Source WAFs Based on OWASP CRS Paranoia Levels: A Systematic Approach for Comparative Measurement †. *Engineering Proceedings*, 115(1). <https://doi.org/10.3390/engproc2025115001>
- Román-Gallego, J. Á., Pérez-Delgado, M. L., Viñuela, M. L., & Vega-Hernández, M. C. (2025). Artificial Intelligence Web Application Firewall for advanced detection of web injection attacks. *Expert Systems*, 42(1). <https://doi.org/10.1111/exsy.13505>
- Şakar, T., & Emekci, H. (2025). Maximizing RAG efficiency: A comparative analysis of RAG methods. *Natural Language Processing*, 31(1), 1–25. <https://doi.org/10.1017/nlp.2024.53>
- Saleh, A. O. M., Tur, G., & Saygin, Y. (2025). SG-RAG MOT: SubGraph Retrieval Augmented Generation with Merging and Ordering Triplets for Knowledge Graph Multi-Hop Question Answering. *Machine Learning and Knowledge Extraction*, 7(3). <https://doi.org/10.3390/make7030074>

- Scano, C., Floris, G., Montaruli, B., Demetrio, L., Valenza, A., Compagna, L., Ariu, D., Piras, L., Balzarotti, D., & Biggio, B. (2024). *ModSec-Learn: Boosting ModSecurity with Machine Learning*. https://doi.org/10.1007/978-3-031-76459-2_3
- Shaon, M. S. H., & Akter, M. S. (2025). Modern Approaches to Software Vulnerability Detection: A Survey of Machine Learning, Deep Learning, and Large Language Models. In *Electronics (Switzerland)* (Vol. 14, Number 22). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/electronics14224449>
- Sheikholeslami, S., Ghasemirahni, H., Payberah, A. H., Wang, T., Dowling, J., & Vlassov, V. (2025). Utilizing Large Language Models for Ablation Studies in Machine Learning and Deep Learning. *EuroMLSys 2025 - Proceedings of the 2025 5th Workshop on Machine Learning and Systems*, 230–237. <https://doi.org/10.1145/3721146.3721957>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Tihanyi, N., Ferrag, M. A., Jain, R., Bisztray, T., & Debbah, M. (2024). *CyberMetric: A Benchmark Dataset based on Retrieval-Augmented Generation for Evaluating LLMs in Cybersecurity Knowledge*. <http://arxiv.org/abs/2402.07688>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. <http://arxiv.org/abs/2201.11903>
- Xu, H., Wang, S., Li, N., Wang, K., Zhao, Y., Chen, K., Yu, T., Liu, Y., & Wang, H. (2025). Large Language Models for Cyber Security: A Systematic Literature Review. *ACM Transactions on Software Engineering and Methodology*. <https://doi.org/10.1145/3769676>
- Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). *REACT: SYNERGIZING REASONING AND ACTING IN LANGUAGE MODELS*. <https://doi.org/https://doi.org/10.48550/arXiv.2210.03629>
- Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J., & Cui, B. (2026). Retrieval-Augmented Generation for AI-Generated Content: A Survey. In *Data Science and Engineering*. Springer Nature. <https://doi.org/10.1007/s41019-025-00335-5>
- Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). *Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena*. <http://arxiv.org/abs/2306.05685>

Zhou, X., Cao, S., Sun, X., & Lo, D. (2025). Large Language Model for Vulnerability Detection and Repair: Literature Review and the Road Ahead. In *ACM Transactions on Software Engineering and Methodology* (Vol. 34, Number 5). Association for Computing Machinery. <https://doi.org/10.1145/3708522>