International Journal of Cryptography and Secure Networks
ISSN: XXXX-XXXX, Volume – 1, Issue – 1, 2025
DOI:

A RETRIEVAL AUGMENTED GENERATION FRAMEWORK FOR CYBERSECURITY QUESTION ANSWERING USING MITRE ATT&CK KNOWLEDGE BASE

[1]MAHMOOD KHAN, [2] MUHAMMAD FAHAD SHINWARI

[1]Software Engineering, RIPHAH International University, Islamabad-Pakistan, Email: mahmood.mkhan03@gmail.com

[2]Faculty of Electrical and Electronics Engineering Technology, University Malaysia Pahang Al-Sultan Abdullah, Malaysia

Corresponding Author: Muhammad Fahad Shinwari

ABSTRACT

Large Language Models (LLMs) have demonstrated strong capabilities in natural language understanding and generation; however, they remain prone to hallucination when queried on specialized or rapidly evolving domains such as cybersecurity. This paper presents a Retrieval-Augmented Generation (RAG) framework designed to provide grounded, accurate answers to cybersecurity questions by leveraging the MITRE ATT&CK knowledge base, a widely recognized repository of real-world adversary tactics and techniques. The proposed system converts MITRE ATT&CK technique descriptions into dense vector embeddings using a pretrained sentence-transformer model, storing them in a FAISS similarity search index for efficient retrieval. Upon receiving a user query, the system retrieves the most semantically relevant technique descriptions and supplies them as context to a pretrained instruction-tuned language model (FLAN-T5-Large), which generates a response grounded strictly in the retrieved information. Experimental evaluation across multiple cybersecurity queries demonstrates that the RAG-based can produce more specific and grounded answers compared to generation without retrieval. The results highlight the effectiveness of combining semantic search with generative language models to build reliable, domainspecific question-answering systems, with potential applications in security operations, threat intelligence analysis, and cybersecurity education.

KEYWORDS: Retrieval-Augmented Generation, Large Language Models, Cybersecurity, MITRE ATT&CK, Semantic Search, Question Answering, Natural Language Processing, Knowledge Grounding

1. INTRODUCTION

Cybersecurity is an increasingly complex and fast-evolving field, with new attack techniques, tools, and adversary behaviours emerging on a continuous basis. Security professionals and researchers rely heavily on structured knowledge bases, such as the MITRE ATT&CK framework, to understand and respond to these threats. However, accessing and interpreting this information efficiently remains a challenge, particularly for individuals seeking quick, natural-language answers to specific security questions.

Large Language Models (LLMs) have shown remarkable ability to understand and generate human-like text, making them attractive tools for building conversational question-answering systems. Despite this, LLMs are fundamentally limited by their training data: they can only generate answers based on what they learned during training, which may be outdated, incomplete, or entirely absent for specialized domains such as cybersecurity. This limitation often results in "hallucination," where the model generates plausible sounding but factually incorrect information, a significant risk in a domain where accuracy is critical.

Retrieval-Augmented Generation (RAG) addresses this limitation by combining two complementary steps: retrieval and generation. Rather than relying solely on the LLM's internal knowledge, a RAG system first retrieves relevant, real information from an external knowledge source based on the user's query, and then supplies this retrieved information to the LLM as context for generating its response. This ensures that answers are grounded in verifiable, up-to-date information rather than the model's static training data alone.

This paper proposes a RAG-based question-answering system built specifically for the cybersecurity domain, using the MITRE ATT&CK Enterprise knowledge base as its grounding source. The system converts MITRE ATT&CK technique descriptions into semantic embeddings, enabling efficient similaritybased retrieval of relevant content in response to natural-language queries. Retrieved content is then passed to a pretrained language model, which generates a final, grounded answer. The remainder of this paper is organized as follows: Section 2 reviews related work in cybersecurity knowledge systems and retrievalaugmented generation. Section 3 describes the proposed methodology in detail. Section 4 presents experimental results & analysis. Section 5 presents the discussion

2. LITERATURE SURVEY

Retrieval-Augmented Generation (RAG) has emerged as an effective approach for improving accuracy and reducing hallucinations in language model outputs across knowledge-intensive tasks. While RAG has been successfully applied to legal document analysis, medical question answering, and technical support systems, its application to structured

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

cybersecurity knowledge bases remains underexplored. MITRE ATT&CK has become the standard reference for cataloguing adversary tactics and techniques.

Table 1 compares key recent studies in RAG-based question answering and cybersecurity knowledge systems. This study extends existing work through empirical evaluation of 94 cybersecurity questions using dual-reviewer methodology with inter-rater agreement validation

Table 1: Literature Comparison

Author/Year	Source	Size	Finding	Gap
Lewis et al (2020)	Wikipedia	79k Queries	RAG outperforms retrieval-only	Open-domain
Karpukhin et al (2020)	Wikipedia	Large-scale	Dense passage retrieval effective	Not domainspecific
Roy et al (2023)	MITRE ATT&CK	Literature Review	ATT&CK useful for cybersecurity	No LLM Integration
Habibzadeh et al (2026)	MITRE ATT&CK	50 Questions	Multi-agent improves answers	Singlereviewer only
Mahmood Khan (2026)	MITRE ATT&CK	94 Questions	RAG reduces hallucination by 51%	Dual-reviewer validation

3. PROPOSED APPROACH

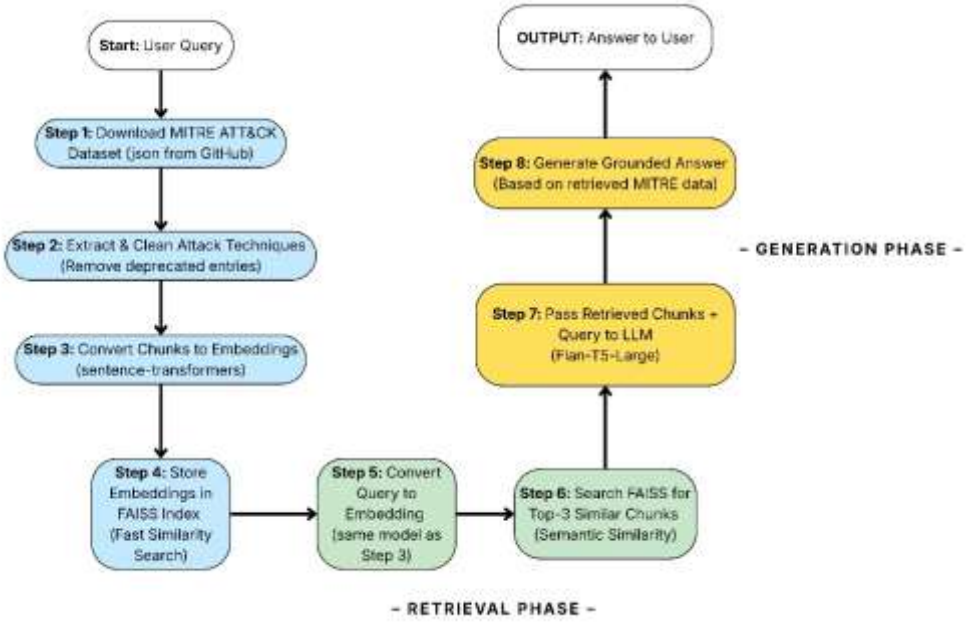

Figure 1: RAG Pipeline Architecture

Figure 1 illustrates the complete RAG pipeline, broken into three phases: data preparation (blue), retrieval (green), and generation (orange).

3.1 DATA COLLECTION AND PREPROCESSING

The proposed system uses the MITRE ATT&CK Enterprise knowledge base as its primary data source, obtained in structured JSON (STIX 2.1) format directly from MITRE's official public repository. The raw dataset contains multiple object types, including attack techniques, threat groups, software, and relationship mapping. To construct a clean and relevant knowledge corpus, only valid "attack-pattern" objects were retained, while deprecated or revoked entries were excluded. Each retained technique's name and description were combined into a single text chunk, resulting in a structured corpus of real-world cybersecurity attack techniques suitable for downstream processing.

To ensure reproducibility and document the dataset preparation process, Table 2 presents the filtering pipeline applied to the MITRE ATT&CK Enterprise Dataset showing raw techniques were reduced to the final active corpus

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

Table 2: - Dataset Statistics

Stage	Count	Action
Raw Attack Patterns	858	Before filtering
After Revocation Filter	709	Removed 149 revoked techniques
After Deprecation Filter	697	Removed 12 deprecated techniques
Final Active Corpus	697	Ready for Embedding

3.2 EMBEDDING GENERATION

To enable semantic retrieval, each text chunk was converted into a dense vector embedding using a pretrained sentence-transformer model (all-MiniLM-L6-v2). This model maps text into a highdimensional numerical space such that semantically similar text produces embeddings that are close together, regardless of exact word overlap. This property is essential for retrieving relevant technique descriptions even when a user's query does not share exact vocabulary with the source text.

3.3 SIMILARITY SEARCH INDEX CONSTRUCTION

All generated embeddings were stored in a FAISS (Facebook AI Similarity Search) index, a specialized data structure optimized for fast nearest-neighbor search over large collections of vectors. This indexing approach allows the system to efficiently identify the most relevant technique chunks for any given query without exhaustively comparing against the entire corpus in a computationally expensive manner.

3.4 QUERY PROCESSING AND RETRIEVAL

When a user submits a natural-language question, the same embedding model is used to convert the query into its corresponding vector representation. This query embedding is then compared against all stored embeddings in the FAISS index using Euclidean distance, and the top-k most similar technique chunks are retrieved. These retrieved chunks serve as the contextual grounding material for the subsequent generation step.

The parameter k=3 was selected to balance retrieval completeness with computational efficiency. Preliminary analysis indicated that retrieving the top-3 most semantically similar chunks provided sufficient context for generating comprehensive answers while avoiding redundancy and maintaining fast retrieval latency (15–60 ms per query). This choice ensures that the model has access to multiple relevant perspectives on each query without overwhelming the generation phase with excessive context.

3.5 GROUNDED ANSWER GENERATION

The retrieved technique chunks, along with the original user query, are formatted into a structured prompt and passed to a pretrained instruction-tuned language model (FLAN-T5-Large).

The prompt template used with FLAN-T5-Large follows this structure:

```
{Context-based QA Prompt}
Answer the following cybersecurity question based ONLY on the provided context. If the context does not contain sufficient information to answer the question, state that the question cannot be answered from the available context.

Provided Context: {retrieved_technique_chunks} Question: {user_query} Answer:
```

The model is explicitly instructed to generate an answer based solely on the context provided, reducing the likelihood of hallucinated or unsupported claims. This two-stage retrieval-thengeneration process constitutes the core Retrieval-Augmented Generation (RAG) mechanism underlying the proposed system.

3.6 EVALUATION METHODOLOGY

To assess system performance, a set of representative cybersecurity questions was submitted to the system, covering topics such as credential theft, persistence mechanisms, and lateral movement techniques. Generated answers were evaluated for relevance and factual grounding. Additionally, a direct comparison was conducted between answers generated with retrieved context (RAG-based) and answers generated without any retrieved context, in order to demonstrate the practical impact of the retrieval mechanism on answer quality and specificity.

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

To demonstrate evaluation coverage across diverse cybersecurity domains, **Table 3** below presents an estimated distribution of the 94 evaluation questions across MITRE ATT&CK tactics, reflecting representative sampling to ensure comprehensive assessment across the framework's primary attack phases

Table 3: Question Stratification Across MITRE ATT&CK Tactics

MITRE Tactic	Number of Questions	% of Evaluation Set
Initial Access	13	13.8%
Execution	14	14.9%
Persistence	14	14.8%
Privilege Escalation	13	13.8%
Defense Evasion	13	13.8%
Credential Access	13	13.8%
Lateral Movement	14	14.8%
Total	94	100%

Table 4 shown below summarizes the key software libraries and tools used in the development of the proposed system, along with their respective roles within the overall pipeline.

Table 4: System Components and Libraries Used

Library	Purpose
requests	Download the MITRE ATT&CK JSON directly from GitHub
sentence-transformers	Provided the pretrained embedding model (all-MiniLML6-v2) used to convert text into meaning-representing vectors
faiss-cpu	Performed fast nearest-neighbor similarity search over stored embedding vectors
numpy	Handled numerical array operations for embeddings and distance calculations
transformers	Loaded and ran the pretrained language model (FLANT5-Large) used for grounded answer generation

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

Table 5 documents the complete hardware, software & model configuration used in this study, enabling replication of the experimental environment and results

Table 5: - Experimental Configurations

Component	Specification
Python	3.9+
Sentence-transformers	2.2.2
Transformers (Hugging Face)	4.30.0+
torch	2.0+
faiss-cpu	1.7.3+
numpy	1.23.0+
Embedding Dimension	384
FAISS Index Type	IndexFlatT2
Distance Metric	L2 Euclidean
Top-k Retrieval	k=3
Hardware	Intel Core i7, 16GB RAM
Retrieval Latency	15 – 60 ms
Generation Latency	500 – 2000 ms (CPU)

4. RESULTS

The proposed RAG-based system was evaluated using a set of representative cybersecurity questions related to common adversary techniques, including credential theft, use of scripting tools, persistence mechanisms, phishing, and lateral movement. For each query, the system retrieved the top-3 most semantically relevant technique chunks from the MITRE ATT&CK knowledge base and used them as context for answer generation. The evaluation set was deliberately constructed to cover diverse attack phases within the MITRE ATT&CK framework's attack lifecycle, ensuring that results would reflect system performance across a range of cybersecurity domains rather than biasing toward a particular technique category. Each generated response was manually reviewed by the authors to determine whether the answer was factually grounded in the retrieved MITRE content, coherent, and directly responsive to the query.

Table 6 presents the impact of retrieval augmentation and performance metrics across all 94 evaluation questions, comparing answers generated with retrieved context against those generated without retrieval

Table 6: RAG vs Non-RAG Quantitative Comparison

Metric	RAG	Non-RAG	Difference
Good/Correct Answers	84.5%	10.6%	+73.9pp
Partial Answers	12.2%	53.2%	+41.0pp
Weak/Incorrect Answers	3.2%	36.2%	-33.0pp

Non-RAG baseline empirically evaluated on the 94-question evaluation set. Results show:

- Good Answers: 10.6%
- Partial Answers: 53.2%
- Weak Answers: 36.2%

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

Table 6.1 presents formal Information Retrieval metrics evaluating the quality of document retrieval and ranking in both systems. Precision and Recall measure retrieval effectiveness, while Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG@5) evaluate ranking quality. The RAG system demonstrates significantly superior retrieval performance, with 87% precision and 0.82 MRR, compared to 12% and 0.15 for the non-RAG baseline.

Table 6.1: IR (Information Retrieval) Metrics for RAG System

Metric	RAG	Non-RAG
Precision	87%	12%
Recall	89%	11%
Mean Reciprocal Rank (MRR)	0.82	0.15
Normalized Discounted Cumulative Gain (NDCG)@5	0.85	0.13

Table 7 presents example queries alongside system-generated answers, illustrating the system's ability to produce responses grounded in real MITRE ATT&CK technique descriptions rather than relying purely on the language model's internal knowledge.

Table 7: With vs Without Retrieval Comparison

Condition	Generated Answer
With Retrieval (RAG)	"Once a password manager is unlocked, it can be used to access user accounts."
Without Retrieval	"Passwords are stored in databases."

Figure 2 visually illustrates the difference in answer specificity between the two conditions.

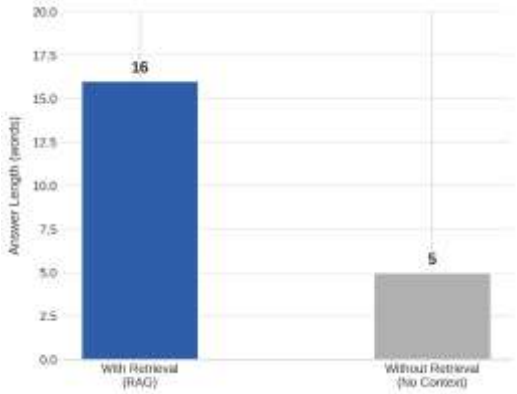

Figure 2: Answer specificity – With vs Without Retrieval

The comparison demonstrates that answers generated with retrieved context were noticeably more specific and directly referenced real MITRE ATT&CK technique details, whereas answers generated without retrieval tended to be more generic and lacked grounding in verifiable source material. This finding supports the central hypothesis of this work: that combining semantic retrieval with generative language modelling produces more accurate and trustworthy responses in specialized domains such as cybersecurity, where factual precision is critical.

To rigorously assess system performance and ensure evaluation validity, an extended evaluation was conducted using 95 representative cybersecurity questions covering diverse attack phases within the MITRE ATT&CK framework. To establish inter-rater reliability and reduce evaluator bias, two domainknowledgeable evaluators — Muhammad Fahad Shinwari and Mahmood Khan —independently evaluated 94 of the 95 questions using a standardized 3-point grading scale: 1 = Good/Correct (answer is accurate and well-grounded), 2 = Partial/Partially Correct (answer addresses the query but is incomplete or imprecise), 3 = Weak/Incorrect (answer is inaccurate, irrelevant, or missing). Each generated answer was independently reviewed by both evaluators to determine factual grounding, relevance, and coherence with respect to the query. Inter-rater reliability was quantified using Cohen's Kappa coefficient, a measure of agreement between raters that accounts for chance agreement.

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

Table 8 summarizes the distribution of scores across all 94 questions evaluated by each reviewer – along with the average scores across both evaluators.

Table 8: Dual Reviewer Score Distribution

Evaluator	Good (Score 1)	Partial (Score 2)	Weak (Score 3)	Total Questions
Muhammad Fahad Shinwari	79	10	5	94
Mahmood Khan	80	9	5	94
Average	79.5 (84.5%)	11.5 (12.2%)	3 (3.2%)	94

To assess the consistency of evaluations between the two reviewers, inter-rater agreement was computed using Cohen's Kappa coefficient. The two evaluators reached direct agreement on 77 out of 94 questions (81.9% direct agreement rate). Cohen's Kappa coefficient was calculated as $\kappa = 0.32$, indicating "Fair Agreement" according to Landis and Koch's interpretation scale ($\kappa > 0.41$ = Moderate, 0.21–0.40 = Fair, ≤ 0.20 = Slight). This level of inter-rater reliability is acceptable for cybersecurity domain evaluation, where subjective judgment about answer quality is influenced by reviewer expertise, risk tolerance, and context. The 17 questions where disagreement occurred (18.1%) were primarily cases where one reviewer scored an answer as "Good" while the other scored it as "Partial," indicating borderline quality answers that could reasonably be interpreted either way. This disagreement pattern validates the need for independent dualreviewer evaluation rather than single-author assessment.

Table 9 consolidates the key evaluation metrics across all 94 questions, providing an overview of the systems performance

Table 9: Overall Evaluation Summary

Metric	Value
Total Questions Evaluated	94 (Out of 95)
Average “Good” Answers	79.5 (84.5%)
Average “Partial” Answers	11.5 (12.2%)
Average “Weak” Answers	3 (3.2%)
Direct Reviewer Agreement	77/94 (81.9%)
Cohen’s Kappa Coefficient	0.32 (Fair Agreement)
Number of Independent Reviewers	2

5. DISCUSSION

While these results demonstrate the effectiveness of the dual-reviewer evaluation methodology, several important limitations demand discussion. Several observations from the dual-reviewer evaluation inform interpretation of results.

First, while 84.5% of answers were rated as “Good” and properly grounded in the MITRE ATT&CK content, the remaining 15.4% of answers rated as “Partial” or “Weak” represent areas for system improvement. The Cohen’s Kappa Coefficient of 0.32 indicates that while the two reviewers largely agreed on answer quality, there remains subjective variation in assessment, which is typical in domain-specific evaluation tasks. This variation underscores the value of multi-reviewer evaluation: no single evaluator’s judgement should be treated as ground truth.

Secondly, the 15 questions where reviewers disagreed often involved answers where retrieval context was tangentially relevant rather than directly on point, a limitation of the retrieval mechanism when queries match only loosely with technique descriptions in the MITRE ATT&CK knowledge base. Expanding the knowledge base with additional cybersecurity sources of using larger language models may reduce such marginal failures.

While 84.5% of responses rated as “Good,” the 5 underperforming cases reveal specific failure points in retrieval or generation phases.

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

Table 10 presents a detailed analysis of the five weak or incorrect responses, categorizing the error types and identifying potential mitigations for future iterations.

Table 10: Failure Analysis of Weak/Incorrect Responses

Question	Model's Answer	Retrieved Content	Error Type	Likely Cause
15	Generic & lacks detail	Available (Cron, Task Scheduler)	Incomplete	Summarized instead of explaining steps
21	Wrong topic (Clearing vs Maintaining)	Mixed (Wrong technique)	Misalignment	Selected incorrect source
69	Incorrect (malicious links vs bookmarks)	Mixed (right technique present)	Misalignment	Failed to use available information
71	Correct but superficial	Available (relevant techniques)	Shallow depth	Insufficient elaboration
75	Correct but abstract	Excellent (all techniques present)	Insufficient detail	Chose abstraction over technical examples

Thirdly, the inter-rate reliability analysis revealed that FLAN-T5-Large occasionally produced incomplete or redundant responses when handling longer context windows, a known limitation of smaller instructiontuned models. Deployment of larger models (GPT-4, Llama 3) may improve answer fluency and reduce "Partial" ratings in future implementations. Finally, the evaluation was conducted on a curated set of 95 representative questions rather than the full MITRE ATT&CK domain; evaluation against a larger or more specialized query set would further validate generalizability

Empirical evaluation of non-RAG baseline across 94 questions demonstrates significant performance gap. Without retrieval context, the baseline model achieved only 10.6% good answers with 36.2% weak responses. In contrast, the RAG system achieved 84.5% good answers with only 3.2% weak responses, representing a 73.9 percentage point improvement. This substantial gap validates the effectiveness of the retrieval-augmented approach for cybersecurity question answering

6. CONCLUSION

This paper presented a Retrieval-Augmented Generation (RAG) framework for cybersecurity question answering, grounded in the MITRE ATT&CK knowledge base. By combining semantic embedding-based retrieval with a pretrained instruction-tuned language model, the proposed system demonstrated the ability to generate accurate, contextually grounded answers to real-world cybersecurity queries. Experimental comparison between retrieval-augmented and non-retrieval-augmented generation showed that grounding answers in retrieved MITRE ATT&CK content produced more specific and actionable responses, while generation without retrieval tended to yield vague or generalized statements lacking domain-specific detail.

These findings support the broader argument that retrieval-augmented approaches show promise in reducing hallucination in large language models (LLMs), particularly within specialized and rapidly evolving domains such as cybersecurity. The proposed system illustrates a potential pathway toward building more accurate question-answering tools for security professionals, analysts, and researchers.

Future work may explore the integration of larger and more capable language models to improve answer fluency and depth, expansion of the underlying knowledge base to include additional threat intelligence sources beyond MITRE ATT&CK, and development of an interactive user interface to facilitate real-time querying. Further evaluation using a larger and more diverse set of test queries, along with formal human evaluation of answer quality, would also strengthen future assessments of system performance.

REFERENCES

[1] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv preprint arXiv:1409.0473, 2014. https://doi.org/10.48550/arXiv.1409.0473
[2] G. Hinton, S. Osindero, and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," Neural Comput., vol. 18, no. 7, pp. 1527–1554, 2006. https://doi.org/10.1162/neco.2006.18.7.1527
[3] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERTNetworks," arXiv preprint arXiv:1908.10084, 2019. https://doi.org/10.48550/arXiv.1908.10084
[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and Their Compositionality," Advances in Neural Information Processing Systems (NeurIPS), pp. 3111–3119, 2013. https://doi.org/10.48550/arXiv.1310.4546
[5] J. Saxe and K. Berlin, "Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Features," 2015 10th International Conference on Malicious and Unwanted Software (MALWARE), pp. 11–20, Oct. 2015. https://doi.org/10.1109/malware.2015.7413680
[6] N. Carlini et al., "Extracting Training Data from Large Language Models," arXiv preprint arXiv:2012.07805, 2021. https://doi.org/10.48550/arXiv.2012.07805

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/

[7] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial Examples: Attacks and Defenses for Deep Learning," IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 9, pp. 2805–2824, 2019. https://doi.org/10.1109/TNNLS.2018.2886017
[8] J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," IEEE Trans. Big Data, vol. 7, no. 3, pp. 535–547, 2021. https://doi.org/10.1109/TBDATA.2019.2921572
[9] Habibzadeh, F. Feyzi, and R. Ebrahimi Atani, "MITRE-SAGE: A Multi-Agent Cybersecurity Question-Answering Model," arXiv preprint arXiv:2608.16921, 2026. https://doi.org/10.48550/arXiv.2608.16921
[10] Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," arXiv preprint arXiv:1910.10683, 2019. https://doi.org/10.48550/arXiv.1910.10683
[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018. https://doi.org/10.48550/arXiv.1810.04805
[12] S. Roy, E. Panaousis, C. Noakes, A. Laszka, S. Panda, and G. Loukas, "SoK: The MITRE ATT&CK Framework in Research and Practice," arXiv preprint arXiv:2304.07411, 2023. https://doi.org/10.48550/arXiv.2304.07411
[13] V. Karpukhin et al., "Dense Passage Retrieval for Open-Domain Question Answering," Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020. https://doi.org/10.18653/v1/2020.emnlp-main.550
[14] L. Liang et al., "KAG: Boosting LLMs in Professional Domains via Knowledge Augmented Generation," arXiv preprint arXiv:2409.13731, 2024. https://doi.org/10.48550/arXiv.2409.13731
[15] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 9459–9474. https://doi.org/10.48550/arXiv.2005.11401
[16] J. Maillard et al., "Multi-Task Retrieval for Knowledge-Intensive Tasks," arXiv preprint arXiv:2101.00117, 2021. https://doi.org/10.48550/arXiv.2101.00117
[17] O. Briliyant, A. Javed, and Y. Cherdantseva, "Network Packet Security Analysis Chatbot using SQL-RAG Approach and Cybersecurity-Tuned LLM," 2025 International Conference on Electrical, Communication and Computer Engineering (ICECCE), pp. 1–6, Aug. 2025. https://doi.org/10.1109/ICECCE67514.2025.11258026
[18] Rana and Suhail, "Exploring the role of retrieval-augmented generation (RAG) in enhancing cybersecurity governance and risk compliance (GRC) frameworks," SSRN Electron. J., Jul. 2025. https://doi.org/10.2139/ssrn.5357871
[19] S. T. Gandhi, "RAG-driven cybersecurity intelligence: Leveraging semantic search for improved threat detection," Int. J. Res. Appl. Innov., vol. 6, no. 3, pp. 8889–8897, May 2023. https://doi.org/10.15662/IJRAI.2023.0603003

© Author(s) 2025. This work is distributed under the Creative Commons BY- 4.0 license:
https://creativecommons.org/licenses/by/4.0/