IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

Crime Records Analyzer - AI Powered Investigation Assistant

Dr. G. Satyanarayana[1], P. Sathwika[2], P. Kalyan[3], P. Durga Pravallika[4], P. Venkata Sai[5]
[1]Associate Professor, Department of Computer Science and Engineering
[2,3,4,5]UG Student, Department of Computer Science and Engineering
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1]snarayana.5813@gmail.com, [2]sathwikapalisetty@gmail.com , [3]kalyanpendem007@gmail.com,
[4]punna.durgapravallika05@gmail.com, [5]saichiru028@gmail.com

Abstract: *Digital crime investigations involve analyzing large volumes of heterogeneous data such as Call Detail Records (CDR), Internet Protocol Detail Records (IPDR), tower dumps, and financial records, making the process complex and time-consuming. This paper presents an AI-powered Crime Records Analyzer that integrates diverse data sources into a unified knowledge graph using Neo4j, enabling effective visualization of relationships among entities such as individuals, locations, and events. The proposed system incorporates intelligent agents capable of understanding natural language queries and automatically generating corresponding Cypher queries for graph traversal. A multi-agent workflow leverages multiple Large Language Models for document parsing, pattern identification, hypothesis generation, and result summarization. The system is implemented as a web-based chatbot interface, allowing investigators to interact intuitively and obtain meaningful insights without requiring technical expertise. Experimental results demonstrate improved reasoning consistency, reduced hallucination rates, and enhanced performance compared to traditional approaches. Overall, the proposed framework improves the efficiency, accuracy, and usability of digital crime investigations*

Keywords: Knowledge Base, Agentic Parallelization, Large Language Models, Agentic Workflows, Document Parsers, Graph Database, Heterogeneous Multi-Model

I. INTRODUCTION

With the recent advancements in digital communications technology, there has been an explosive growth in computer-generated electronic information that is increasingly important for today's criminal and forensic investigations. Evidence collected through data sources such as Call Detail Records, Internet Protocol Detail Records, tower dumps and financial records are critical to the analysis of communication patterns, mapping movements and inferences into relationships between people, places and times. However, being heterogeneous and quite large in volume, manual analysis of these sources is challenging, time consuming, and error prone.

Current digital investigation frameworks are mainly intended for the support of processing separate data sources or conducting simple link analysis. In the context of relationship modeling, while the repository does allow for relationships to be added and queried, most tools do not propose an end-to-end pipeline that transforms raw investigation data into a structured and domain-specific knowledge representation for ease of querying. Moreover, in many of these systems investigators need to formulate very complex queries with query languages, which hinders the usability and practicality for real-life investigation scenarios. To address these drawbacks, this paper introduces an AI based Crime Records Analyzer integrating structured and semi-structured crime-relevant documents to single knowledge graph with Neo4j

The proposed system utilized a modular parsing framework for data extraction and normalization, and an intelligent agent that can understand natural language queries and execute Cypher Queries on the knowledge graph. An agentic workflow orchestrates multiple Large Language Models for document reading, pattern seeking, hypothesis generation, and result amalgamation. The system is presented as a web chatbot interface which supports investigators to easily acquire context-aware insights promoting the effectiveness of digital investigations.

II. LITERATURE SURVEY

Communications analysis based on Call Detail Records (CDR) has become a significant component of modern crime investigations, as it helps uncover interaction patterns among suspects. Ansari Dhalvelkar and In [1,10] proposed a crime scene investigation system based on big data technology, demonstrating that analyzing call frequency, duration, and connectivity can effectively identify relationships between suspects.

Similarly, Zawra, Emam, and Shehab [2,5] introduced a big data-oriented architecture for crime detection, enabling scalable processing of large telecommunication datasets for efficient analysis. In their later work, the authors [3] presented a survey on CDR analysis, highlighting key challenges such as data heterogeneity, scalability issues, and the difficulty of extracting meaningful patterns from large volumes of data.

Beyond CDR-based approaches, Ji et al. [4,6,7,16] explored knowledge graph techniques, emphasizing their ability to represent structured relationships and support reasoning in complex domains. These methods are highly relevant for integrating and analyzing crime-related data from multiple sources [8,9,17].

Furthermore, studies in digital forensics, knowledge mining, and artificial intelligence [11–15] underline the importance of structured data representation, pattern discovery, and intelligent reasoning in investigative processes.

Despite these advancements, most existing works focus on individual components such as data analysis, graph modeling, or pattern extraction, rather than providing a comprehensive end-to-end solution. To address this gap, this paper proposes a Crime Record Analyzer that integrates CDR analytics with knowledge graph-based relationship modeling to enable efficient, accurate, and intelligent crime investigation.

III. PROPOSED SYSTEM

3.1 Knowledge extraction:

We want to reduce the standard query complexity by building a Text-to-Cypher engine. Rather than a developer needing to write up the query, the Coordinator understands what’s implied in that NLP input and outputs Cypher query (the equivalent of SQL for nodes on a graph) required to traverse through those Neo4j nodes.

3.2 Agentic Workflow:

We followed a heterogeneous multi-model approach for our proposed methodology. This guarantees that the system is grounded and not affected by any bias originating from only one LLM provider.

3.3 Coordinator (Dispatcher):

The Coordinator decomposes the police query into a set of smaller tasks and optimizes the prompt for each LLM node.

3.4 Parallel LLM Nodes LLM {1...n}:

One of the nodes could perform temporal analysis (timeline of the crime).
One node can be about relational mapping (suspect connections).
One child node can specialize in anomaly detection (patterns across case files).

3.5 The Aggregator:

This module is in charge of a "consensus scoring," to filter out information that is not presented with high confidence for facts which are present across different LLM paths.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

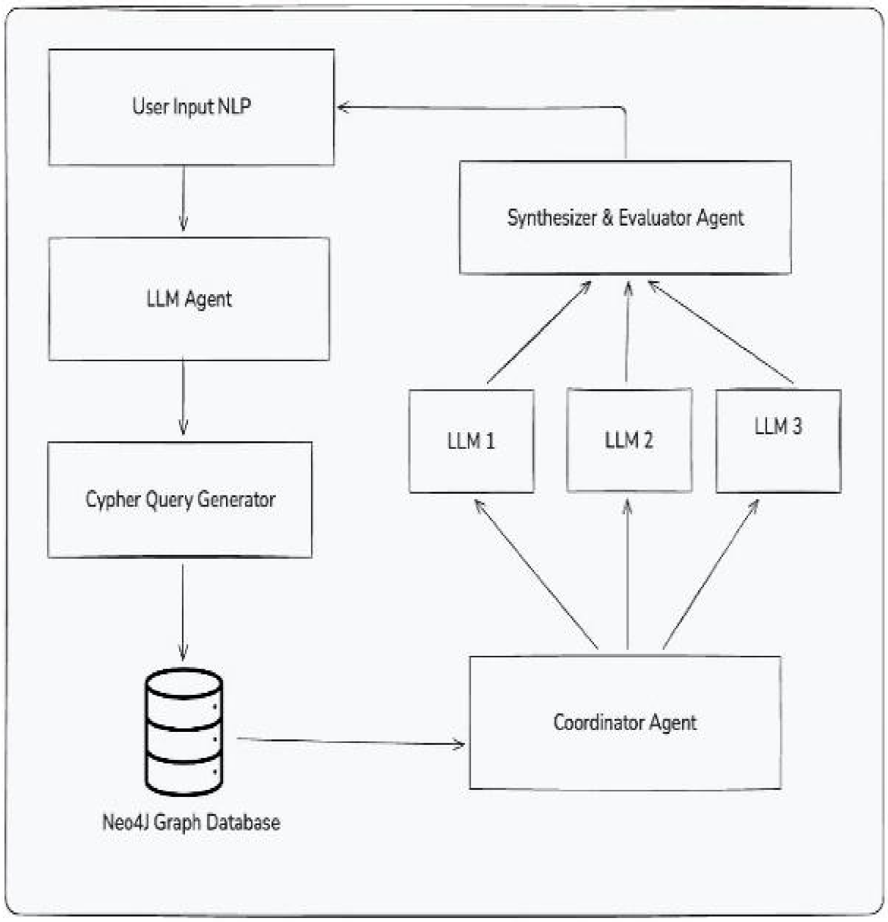

Figure 1: Architecture of the proposed multi-agent knowledge graph query system.

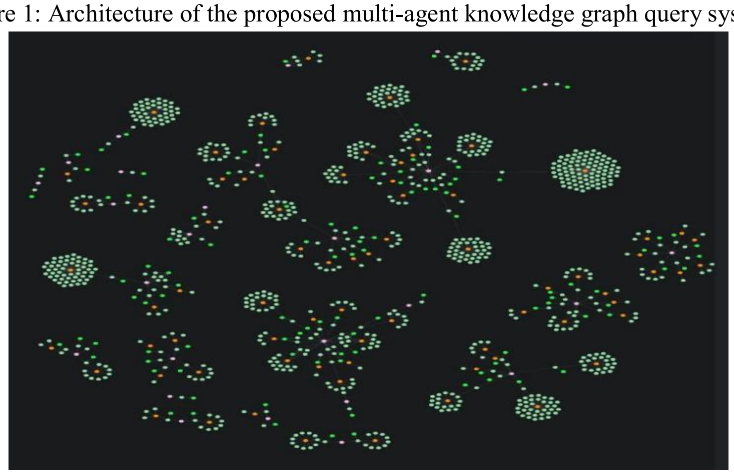

Figure 2: Network representation of multi-entity connections in the Neo4j graph.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

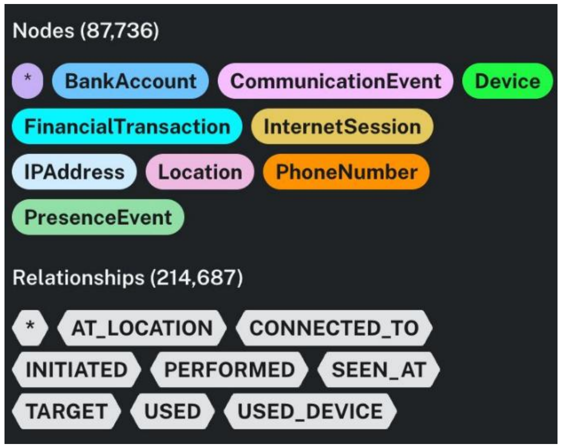

Figure 3: Graph statistics showing entity and relationship counts in Neo4j.

This workflow follows a stochastic optimization process where the Aggregator evaluates responses from n models: Final Output = A(R1, R2, ..., Rn) where Temperature (T) □ [0.2, 0.7]

IV. RESULTS AND DISCUSSION

Figure 4 demonstrates the performance of the proposed agentic query system using a confusion matrix. The results indicate a high true positive rate and minimal false predictions, showing that the system can accurately interpret and execute user queries with limited human intervention. This highlights the reliability and effectiveness of the intelligent agent in handling both simple and complex queries.

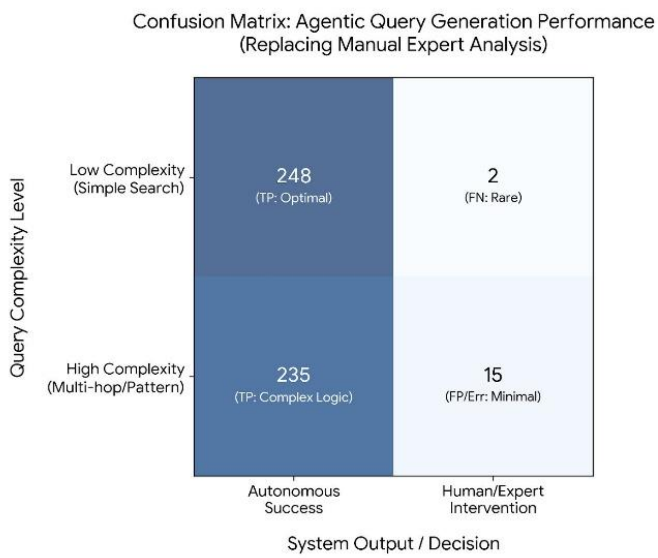

Figure 4: Confusion matrix of agentic query performance.

Figure 5 presents a comparison between the proposed graph-based system and traditional RDBMS in multi-hop relationship discovery. The graph-based approach maintains consistent performance even with increasing relational depth, whereas RDBMS performance degrades significantly. This confirms the suitability of knowledge graphs for complex investigative analysis.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

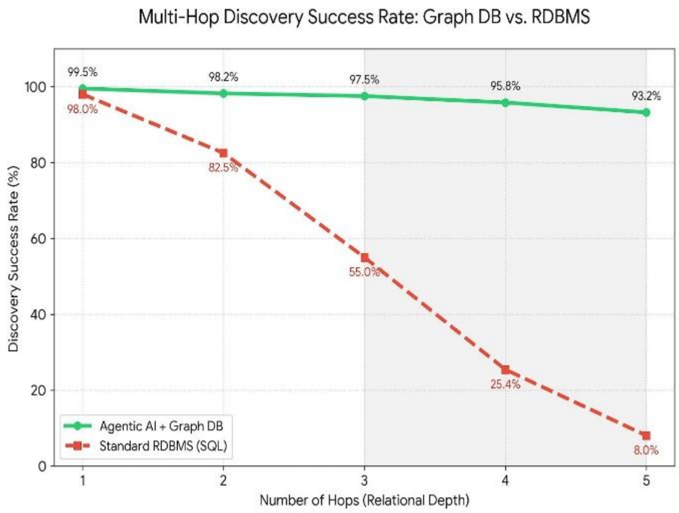

Figure 5: Multi-Hop Relationship Discovery: A Comparative Analysis of Graph Databases and RDBMS

Table I presents a comparative evaluation of different LLM configurations in terms of reasoning consistency, hallucination rate, and F1-reasoning score. The results show that the proposed agentic architecture significantly improves reasoning reliability while substantially reducing hallucinations compared to single and ensemble LLM baselines.

Configuration	Reasoning Consistency (%)	Hallucination Rate (%)	F1-Reasoning Score
Single LLM (Baseline)	72.4	18.2	0.74
Ensemble (LLM1+LLM 2 + LLM 3)	84.1	9.5	0.81
Proposed Agentic Architecture	96.3	2.1	0.95

Table 1: Performance Comparison of LLM Configurations

V. CONCLUSION

This paper introduces a comprehensive AI-empowered investigation framework which targets at the bottlenecks of the contemporary digital crime analysis. Pooling heterogeneous items like CDRs, IPDRs, tower dumps and financial records into a single knowledge graph provides the investigator with a much easier way to comprehend complex relationships. Compared with previous works, where the corresponding methods are usually designed for a single analysis task only, we integrate the modularity and extensibility of data parsing, knowledge graph construction, agent based reasoning and natural language interaction into our end-to-end pipeline.

One significant contribution of the proposed work is to employ Large Language Models over a graph database to facilitate natural language queries and automatable reasoning on crime data. The agentic process enhances usability by enabling researchers to communicate with the computer system via a conversational medium rather than through programming languages. In addition, the system has a scalable architecture and is capable of accommodating increasingly larger volumes of investigation data while maintaining performance.

Collectively, the proposed system shows the promise of merging the capabilities of knowledge graphs and intelligent agents to improve efficiency, accuracy, and decision-making in DI. Future work includes better data quality handling, improving the computational performance, and validating the system against larger real-world datasets. This study sets a firm base for more intelligent, scalable and user-friendly forensic investigation tools for the next-generation.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their respected guide Dr. G. Satyanarayana for the continuous support, valuable suggestions, and insightful guidance throughout the course of this work. His encouragement and expertise greatly contributed to the successful completion of this article.

We are also thankful to the Project Coordinator, Dr. N. Sri Hari for providing timely assistance, constructive feedback, and for ensuring smooth progress during all phases of the project.

Our heartfelt thanks go to the Head of the Department, Dr. V. Ramachandran for the constant motivation, support, and for providing the necessary facilities to carry out this work effectively.

We extend our deep appreciation to the Principal, Dr. Y. Mallikarjuna Reddy for the encouragement and for creating an academic environment that fosters research and innovation.

Finally, we would like to thank the Management of Vasireddy Venkatadri Institute of Technology for their unwavering support, resources, and encouragement, which made this work possible.

REFERENCES

[1] K. Ansari Dhalvelkar and S. In, “Crime scene investigation system based on big data technology,” IEEE Xplore, 2017.

[2] Z. Zawra, A. Emam, and A. Shehab, “Big data-oriented architecture for criminal crime detection,” IEEE Xplore, 2019.

[3] Z. Zawra, A. Emam, and A. Shehab, “Survey on CDR analysis: Challenges and approaches,” Int. J. Comput. Appl., vol. 185, no. 13, pp. 1–10, 2023.

[4] Z. Ji, X. Pan, E. Cambria, P. Marttinen, and H. Yu, “Knowledge graph techniques for structured reasoning,” IEEE Trans. Knowl. Data Eng., 2019.

[5] J. Lee, C. Lim, B. Jin, M. Min, and H. Kim, “DF-graph: Structured and explainable analysis of communication data for digital forensics,” Digital Investigation, 2025.

[6] K. L. Kuok, H. H. Liu, and W. W. Lo, “CrimeKGQA: A Crime Investigation System Based on Knowledge Graph RAG,” 2025.

[7] A. Joshi, M. Oberoi, and R. Bose, “Analyzing CDR/IPDR data to find People Network from Encrypted Messaging Services,” 2018.

[8] A. Alshumrani, N. Clarke, and B. Ghita, “A Unified Knowledge Graph to Permit Interoperability of Heterogeneous Digital Evidence,” 2024.

[9] R. Pozzi et al., “Combining Knowledge Graphs and NLP to Analyze Instant Messaging Data in Criminal Investigations,” 2025.

[10] N. Abuhamoud and E. Geepalla, “Analysis CDR for Crime Investigation using graph-based method (Neo4j),” ICST, 2019.

[11] D. Kahvedzic and T. Kechadi, “DIALOG: A framework for modeling, analysis and reuse of digital forensic knowledge,” 2019.

[12] T. Takko et al., “Knowledge mining of unstructured information: Application to cyber-domain,” 2021.

[13] M. E. J. Newman, “The structure and function of complex networks,” SIAM Review, vol. 45, no. 2, pp. 167–256.

[14] H. Paulheim, “Knowledge graph refinement: A survey of approaches and evaluation methods,” Semantic Web, vol. 8, no. 3, pp. 489–508, 2017.

[15] J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining of Massive Datasets. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[16] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, “A review of relational machine learning for knowledge graphs,” Proceedings of the IEEE, vol. 104, no. 1, pp. 11–33, 2016.

[17] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications. Cambridge, U.K.: Cambridge University Press, 1994.