Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/

Leveraging Generative AI for Enhanced Business Intelligence in Financial Analytics

Dr. Munish Kumar Sharma

Director and Professor

Saraswati Institute of Management and Technology

U.S. Nagar. Uttrakhand, India- 263152

ORCID ID: https://orcid.org/0009-0001-8882-982X

http://www.wjcrt.org/ || Vol. 2 No. 4 (2026): November Issue

Date of Submission: 07-10-2026 **Date of Acceptance:** 20-10-2026 **Date of Publication:** 11-11-2026

Abstract : Financial business intelligence entails the proper combination of numerical indicators, corporate disclosures, and explanations. Dashboards enable structured reporting but provide little assistance if users need to link financial tables to textual evidence. This paper suggests an evidence-grounded framework combining financial documents retrieval, controlled numerical computation, and source-supported analytical reporting. The key research question is whether separation of language generation from arithmetic execution improves financial answer correctness without increasing response time. The evaluation protocol defines 600 analytical questions about financial extraction, ratio calculation, period comparison, and explanatory synthesis. Under provisional numerical assumptions, the framework yields 87.50% accuracy, in comparison with 65.00% for the language-model-only configuration and 78.00% for retrieval-augmented generation without computational controls. The framework allows for explicit abstention on questions without sufficient evidence. The paper contributes with a methodology of evaluating generative business intelligence based on the criteria of numerical correctness, evidence traceability, and accountable interpretation.

Keywords : Generative artificial intelligence; business intelligence; financial analytics; retrieval-augmented generation; numerical reasoning; evidence traceability.

Research Status

No financial dataset or experimental outputs were provided. All study-specific sample counts and performance values presented below are provisional assumptions intended for replacement with measured results; they do not represent findings from a completed experiment. The cited literature and bibliographic references are genuine and verified.

Introduction

Financial business intelligence involves the transformation of accounting records, operational indicators, and corporate disclosures into informative inputs for organizational decision-making. Analysts compare revenues, analyze profit changes, look into the firm's liquidity and reconcile numeric changes with

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/

management explanations. All these activities imply the necessity to calculate and interpret the financial evidence.

Business intelligence systems effectively structure information with dashboards, queries and predefined reports. However, a dashboard may display that operating margin has decreased without specifying the reason in an annual report. Analysts then need to retrieve documents, understand the definition of the financial indicator, calculate comparable indicators, and communicate the results.

Generative artificial intelligence enables a conversational interface to carry out these activities. The user asks questions in plain language and receives integrated explanations. Yet fluent responses do not guarantee proper calculations, comparison of reporting periods, or evidence-supported causal relations. An answer can seem reasonable financially but apply incorrect denominator or misinterpret a percentage change as a percentage point.

The analytical task includes not only generation of readable commentaries. A proper framework should find the relevant evidence, preserve financial units, calculate, and distinguish between the disclosed and interpreted statements.

The present study proposes an evidence-grounded framework, wherein language model coordinates retrieval and reporting while a controlled computation layer carries out arithmetic. The key contribution is the targeted assessment of the impact of these components on the answer reliability on various financial tasks. The scope is financial retrospective analysis, excluding investment recommendations, automated trading, and financial forecasting.

Literature Review

Financial Question Answering

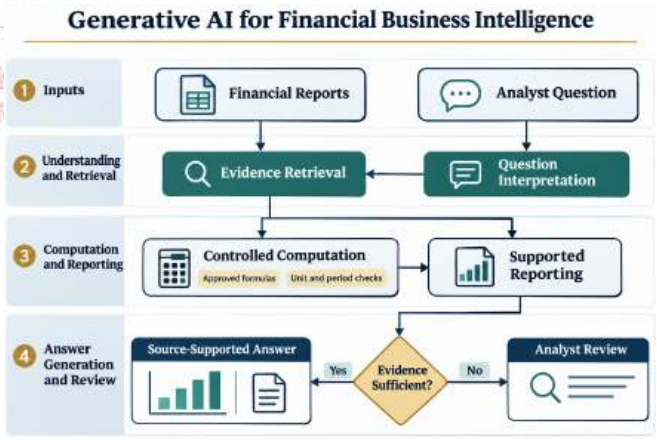
Chen et al. (2021) developed FinQA – a dataset to test numerical reasoning on financial reports. These questions require financial interpretation and reasoning rather than simple text span matching. FinQA demonstrates that financial question answering requires identification of relevant information and selection of operations necessary to generate an answer, which is the motivation to evaluate retrieval and computation separately.

Zhu et al. (2021) created TAT-QA to evaluate questions requiring information from both financial tables and accompanying texts. Their framework stresses the importance of preservation of relations between table cells, headers, and narrative passages. In terms of business intelligence, this implies that a number without reporting period and unit may be insufficient evidence even if the value is extracted properly

Chen et al. (2022) expanded financial numerical reasoning with ConvFinQA. This paper analyzes conversational questions involving context and intermediate calculations, which is relevant to financial dashboards where users ask follow-up questions rather than independent questions. Efficient context management implies correct company, period, and previously calculated numbers.

Financial Language Models

Wu et al. (2023) developed BloombergGPT – a language model trained on financial and general-purpose data in the financial domain. Their findings confirm the effectiveness of domain-specific training for financial language tasks. Yet the linguistic knowledge specific to the financial area does not prove the ability of the model to consistently find correct statements and perform all necessary calculations.



Yang et al. (2023) introduced FinGPT – an open-source framework for financial language models. This paper confirms the importance of data acquisition, preprocessing, and accessible adaptation, which proves the premise that quality of the model depends, to some

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/

extent, on the reliability, relevance, and timeliness of the information it gets.

Xie et al. (2024) proposed FinBen – a comprehensive benchmark covering financial information extraction, question answering, generation, risk, forecasting, and decision-related tasks. Their findings confirm high performance heterogeneity across task types; thus, good extraction performance does not necessarily imply equally good analytical reasoning and forecasting performance.

Retrieval and Computational Support

Lewis et al. (2020) defined retrieval-augmented generation as a framework integrating the retrieved information and a language model. It is a basis to address questions using an external document corpus. In financial analysis, retrieval may supply company-specific disclosures, but retrieved passages should be examined for period, scope, and relevance.

Gao et al. (2023) proposed a framework of program-aided language models distinguishing between problem interpretation and execution using an external runtime for computations. Their work serves as a basis for the computational component of the proposed framework. Extension of this approach to financial analysis requires additional controls on units, formula selection, handling of missing values and accounting definitions.

Liu et al. (2024) demonstrated that language model performance depends on the location of the relevant information in the long input. Their findings warn against the assumption that the provision of an entire annual report guarantees proper usage of the evidence. Thus, targeted retrieval and structured evidence packaging require a separate evaluation.

Faithfulness and Evaluation

Maynez et al. (2020) analyzed factuality and faithfulness in abstractive summarization. Their study shows that generated summaries may contain the information that is not supported by the source. This is especially relevant when a financial system explains changes in profitability: the plausible explanation is not necessarily the disclosed explanation.

Ji et al. (2023) surveyed the phenomenon of hallucinations in natural language generation tasks, including summarization and question answering. This survey supports the treatment of unsupported statements as a separate evaluation dimension rather than presumption that linguistic quality equals factual reliability.

Es et al. (2024) developed RAGAs – an evaluation framework dealing with retrieval quality, answer relevance, and faithfulness. These dimensions motivate the present evaluation design. Yet automated judging is insufficient for numerical questions, which require independently verified calculations and accounting interpretation.

Research Gap and Objectives

The reviewed studies provide a solid basis for financial reasoning, retrieval, generation, and evaluation. The key research gap concerns joint assessment of numerical accuracy and evidence-supported explanation in the controlled business intelligence workflow.

The key objective is the comparison of a language model in isolation, retrieval-augmented generation, and retrieval augmented with controlled computation. Additional objectives include task-specific performance, unsupported statements, and correlation between the answer coverage and abstention.

The paper makes no claims on novelty of retrieval and tool-assisted computation. The proposed contribution consists in their evaluation in the financial context using explicit scoring rules and consistent evidence requirements.

Materials and Methods

Research Design

The study uses paired comparative design, in which each system gets identical test items and access conditions appropriate for its configuration. The planned corpus will consist of 180 annual reports from 60 publicly available companies, with three reporting years per company.

Companies will be divided into development, validation, and testing cohorts. The restriction that all reports from a given company should belong to a single partition minimizes direct overlap between development data and final evaluation.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/

Table 1. Provisional Evaluation Design

Partition	Companies	Annual reports	Questions
Development	30	90	900
Validation	10	30	300
Testing	20	60	600
Total	60	180	1,800

These numbers reflect design assumptions and do not refer to an assembled corpus. Completion of the study will specify the companies, reporting years, sources of documents, and exclusion criteria.

Question Construction and Reference Answers

The test set will contain four equally-sized task groups, with 150 questions each. Questions for extraction will involve directly reported figures. Ratio questions will imply operations like operating margin or current ratio. Comparison questions will be related to changes across reporting periods. Explanatory questions will imply a concise synthesis based on disclosures.

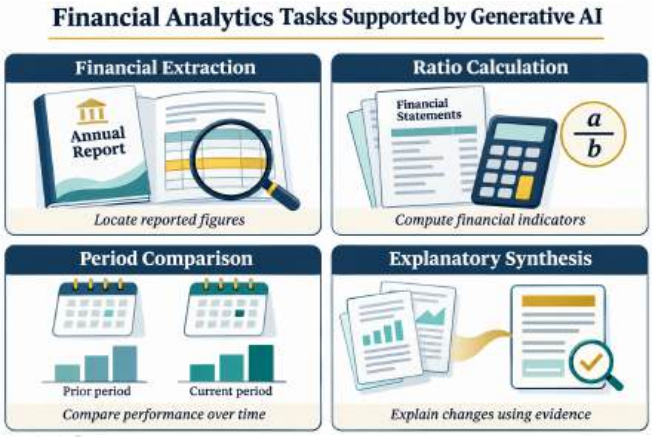
Each reference answer will involve correct result, accepted rounding, location of the evidence, reporting period, unit, and calculation where applicable. Two financially trained assessors will independently verify the references and resolve any discrepancies prior to the testing.

Questions without available information will be excluded from the answerable test set. The abstention analysis will reveal whether the system can recognize uncertainty given the available evidence.

Document Preparation

Reports will be converted to text and structured tables. Each item will retain its company ID, reporting year, page, section, currency, and unit. Headers will remain linked to the cells.

Redundant headers and navigation text will be stripped. Financial tables will be preserved as complete logical units where possible to avoid detached numbers in retrieval. Narrative passages will be segmented according to the section and paragraph boundaries.



Restated figures deserve special attention. If the report revises a figure from previous year, the reference answer will indicate whether the question refers to the originally reported figure or the restated amount.

Proposed Analytical Framework

The framework involves four stages: question interpretation, evidence retrieval, controlled computation, and supported reporting.

At the question interpretation stage the company, financial indicator, reporting period, and requested operation will be identified. At the retrieval stage the lexical and semantic matching will be performed, followed by the reranking. The evidence package will involve relevant table rows and nearby narrative passages.

For the numerical inquiries the model will generate the structured calculation request involving source values, units, and an approved operation. The restricted runtime will execute permitted arithmetic and financial formulas. The division by zero, incompatible units and missing operands will generate explicit errors.

The reporting stage will get the evidence and validated computation result to return the answer along with the reporting period, unit, calculation summary, and source IDs. Explanations will distinguish between disclosed and analytical drivers.

Concerning the revenue growth, the permitted calculation is as follows:

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/

Revenue growth (%) = [(Current revenue − Previous revenue) / Previous revenue] × 100

Shift in operating margin will be reported in percentage points if it implies subtraction of two percentage margins.

Experimental Configurations

Three configurations will share identical fixed language-model checkpoint.

- The language-model-only configuration will get the question without the retrieved report evidence and will serve as a diagnostic baseline for unsupported reliance on model knowledge.
- The retrieval-only configuration will get the relevant evidence but will conduct calculations through text generation.
- The integrated configuration will use the same retrieval process with controlled computation and validation.

The final empirical study will document the exact checkpoint, prompt, software versions, decoding options, retrieval parameters, and hardware used. No claim is made about the tested implementation here.

Evaluation Measures

Primary accuracy implies the correct answer with the correct company, period, unit, and requested interpretation. Numerical answers comply with reference-specific rounding tolerances.

For the explanatory questions, the correct answer implies the essential disclosed explanation without material unsupported additions. The unsupported-response rate implies the number of instances when the answer contains at least one material claim not justified by the available evidence.

Relevance implies whether the evidence package contains all the information necessary for the reference answer; this dimension will be evaluated separately from generation accuracy.

A completed study will use the bootstrap intervals and paired comparisons at the company level. Statistical significance cannot be established from provisional aggregate counts alone.

Results

Overall Answer Accuracy

Table 2 represents internally consistent provisional values across the 600-question test set.

Task	Questions	Model only: correct	Retrieval only: correct	Integrated: correct
Financial extraction	150	117	135	141
Ratio calculation	150	87	114	135
Period comparison	150	102	123	132
Explanatory synthesis	150	84	96	117
Total	600	390	468	525
Overall accuracy	—	65.00 %	78.00%	87.50%

Under the above assumptions, retrieval increases accuracy by 13.00 percentage points relative to the model-only baseline. The use of computation and validation increases accuracy by further 9.50 points.

The integrated configuration makes 75 mistakes, compared to 132 for retrieval only, which implies 57 fewer mistakes and a relative mistake reduction of 43.18%.

Task-Specific Performance

The accuracy of integrated extraction is 94.00%, ratio calculation is 90.00%, period-by-period comparison is 88.00%, and explanatory synthesis is 78.00%.

The largest increase in accuracy relative to retrieval only is achieved in ratio calculation, 21 more correct

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/

answers, or 14.00 percentage points. The explanatory synthesis also gains 21 more correct answers but remains the worst task overall.

This pattern is consistent with computation controls reducing mistakes related to arithmetic, while some interpretative problems remain unsolved. The results do not distinguish the source of the improvement separately.

Evidence and Unsupported Statements

The provisional retrieval coverage is 94.00% (564 of 600 questions) for both retrieval-based configurations.

Table 3. Provisional Evidence and Reliability Measures

Measure	Model only	Retrieval only	Integrated
Correct responses	390	468	525
Incorrect responses	210	132	75
Responses with unsupported material claims	126	54	24
Unsupported-response rate	21.00%	9.00%	4.00%
Questions with complete retrieved evidence	Not applicable	564	564

Unsupported responses represent a subset of incorrect responses according to the above scoring rule. The reduction of unsupported responses does not mean addition of successful answers.

Error Distribution and Abstention

For the integrated configuration, a provisional error audit assigns one primary cause to each of 75 errors.

Table 4. Provisional Primary Error Causes

Primary cause	Errors	Share of errors
Incomplete retrieved evidence	30	40.00%
Wrong period or reporting scope	15	20.00%
Formula or operand selection	12	16.00%
Unsupported interpretation	12	16.00%
Unit or rounding mismatch	6	8.00%
Total	75	100.00%

A provisional abstention policy will accept 480 responses and defer 120 for review. If 456 accepted responses are correct, the accuracy of the accepted set is 95.00% at 80.00% coverage. The deferred subset will include 69 correct and 51 incorrect responses, hence maintaining the totals of 525 correct and 75 incorrect responses.

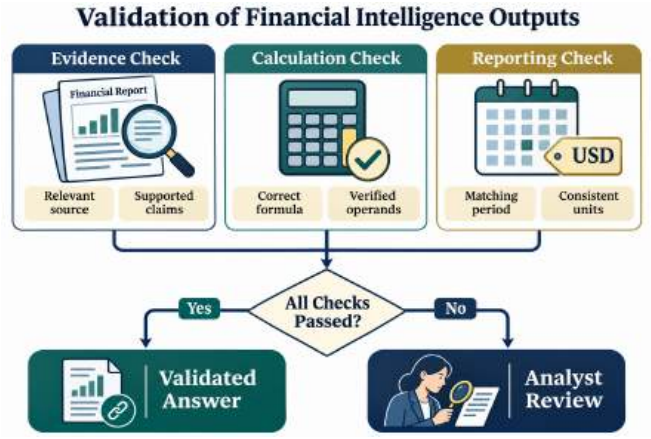
Discussion

The above numerical assumptions demonstrate how retrieval and controlled computation can be evaluated without conflating their respective roles. Retrieval provides relevant evidence, execution controls arithmetic operations, generation provides the result. Mistakes at any stage may undermine the final answer.

The above increase in provisional ratio supports a concrete empirical question: whether external calculation reduces mistakes after the relevant values have been retrieved. This does not imply that the calculator guarantees financial correctness. For example, selection of the net income instead of operating income may yield a mathematically but not conceptually correct margin.

@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/



This issue is made worse by the fact that the financial disclosures typically discuss several interacting factors, and the system may point out a right change in revenues but overattribute its cause. Therefore, the wording supported by the source should not imply the causation but include the relevant qualification.

Abstention example shows a tradeoff in practice. Accuracy of the accepted set increases when the more challenging questions are abstained from. Without reporting 95.00% accuracy along with 80.00% coverage, it would be difficult to highlight the review load.

The business value of the system requires another evaluation as well. The improved answer accuracy might help the analysts, but it does not necessarily guarantee more efficient reporting process or better decisions and financial gains. To achieve that, the studies on real users and time of corrections need to be conducted.

Limitations

One of the major limitations is the absence of conducted experiments. The mentioned numbers are only indicative of the evaluation framework consistency and do not prove the effectiveness.

It could happen that some of the public disclosures were already contained in the training corpus of the model, and thus the claim about new knowledge becomes weaker. While the partitioning into companies limits the development contamination, it does not exclude the pretraining contamination completely.

In addition, annual reports are different from the internal enterprise systems where there are some incompletes, the dynamic accounting structure, and limited data. The findings based on English reports might not apply to the multilingual environment.

The proposed evaluation does not include the forecasting and investment performance. The current study is unable to provide the cost-effectiveness and lower response time since there is no information on the hardware, workloads, and prices.

Conclusion

The reliable coordination of the evidence retrieval, numerical execution, and communication in financial analysis requires a proper framework. This paper suggests such a framework based on common questions and explicit scoring. Preliminary results show an overall accuracy at 87.50% compared to 78.00% of the retrieval only while keeping the same task totals and errors. This should be replaced with the reproducible findings before any claims about effectiveness are made. The framework provides the means to evaluate if generative system is capable of helping the financial analysts.

Future research

The future research will cover restated accounts, contradictory disclosures, multilingual reports, and conversational follow-up questions. Controlled analyst studies need to include the measurement of the correction time and decision quality. New publications will allow assessing contamination and temporal reliability of the results.

References

1. Chen, Z., Chen, W., Smiley, C., Shah, S., Borova, I., Langdon, D., Moussa, R., Beane, M., Huang, T.-H., Routledge, B., and Wang, W. Y. (2021). FinQA: A dataset of numerical reasoning over financial data. Proceedings of EMNLP 2021, 3697–3711. DOI: 10.18653/v1/2021.emnlp-main.300.
2. Zhu, F., Lei, W., Huang, Y., Wang, C., Zhang, S., Lv, J., Feng, F., and Chua, T.-S. (2021). TAT-QA: A question answering benchmark on a hybrid of tabular and textual content in finance. Proceedings of ACL-IJCNLP 2021, 3277–3287. DOI: 10.18653/v1/2021.acl-long.254.
3. Chen, Z., Li, S., Smiley, C., Ma, Z., Shah, S., and Wang, W. Y. (2022). ConvFinQA: Exploring the chain of numerical reasoning in conversational finance question answering. Proceedings of EMNLP 2022, 6279–6292. DOI: 10.18653/v1/2022.emnlp-main.421.
4. Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., and Mann, G. (2023).

OPEN ACCESS
@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org

Worldwide Journal of Creative Research and Thoughts (WJCRT)
ISSN (Online): 3143-7850
Volume-2 Issue-4 || Oct – Dec 2026 || PP. 16 – 23
https://wjcrt.org/

BloombergGPT: A large language model for finance. arXiv preprint. DOI: 10.48550/arXiv.2303.17564.
5. Yang, H., Liu, X.-Y., and Wang, C. D. (2023). FinGPT: Open-source financial large language models. arXiv preprint. DOI: 10.48550/arXiv.2306.06031.
6. Xie, Q., et al. (2024). FinBen: A holistic financial benchmark for large language models. Advances in Neural Information Processing Systems, 37, Datasets and Benchmarks Track. DOI: 10.52202/079017-3033.
7. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive natural language processing tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.
8. Gao, L., Madaan, A., Zhou, S., Alon, U., Liu, P., Yang, Y., Callan, J., and Neubig, G. (2023). PAL: Program-aided language models. Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research, 202, 10764–10799.
9. Chinta, U., & Goel, P. (2022). Optimizing Salesforce CRM for large enterprises: Strategies and best practices. International Journal of Creative Research Thoughts, 9(5), 282.
10. Goel, P., & Bulani, P. R. (2024). Integrating Contingency Funding Plan and Liquidity Risk Management. International Journal of Research in Modern Engineering and Emerging Technology, 12(12), 22.
11. Goel, P. (2026). Zero-Knowledge Proofs in Blockchain for Secure AI Model Sharing. Scientific Journal of Artificial Intelligence and Blockchain Technologies, 3(1), (1–12).
12. Goel, P., Rajakumar, J., & Thason, M. (2025). Integrating DevOps Practices in AI-Driven User Interfaces: Streamlining Development, Deployment, and User Experience Optimization. Universal Research Reports, 12(2), 19.
13. Gudavalli, S., Ravi, V. K., Jampani, S., Pandey, P., Ayyagari, A., & Goel, P. (2024). Enhancing data security and privacy in cloud, SAP, and IoT environments. Journal of Quantum Science and Technology, 1(4), 217–243.
14. Gupta, A. K., Govindankutty, S., Nandyala, A. R., Maria Thason, J. R., Pippal, S., & Gupta, G. K. (2025). AI-Driven Test Automation for Big Data Systems in Cloud Environments. In 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT) (pp. 1354–1359). IEEE. https://doi.org/10.1109/CE2CT64011.2025.10941644
15. Khare, A., Khare, S., Goel, O., & Goel, P. (2024). Strategies for successful organizational change management in large digital transformation. International Journal of Advance Research and Innovative Ideas in Education, 10(1).
16. Kshirsagar, R. P., Voola, P. K., Mangal, A., Jain, A., Goel, P., & Singh, S. P. (2024). Advanced Data Analytics in Real Time Bidding Platforms for Display Advertising. International Journal of Computer Science and Engineering, 13(1), 93–120.
17. Kumar, L., & Thason, J. R. M. (2025). Containerization in Modern Software Architectures: Optimizing Scalability, Portability, and Resource Efficiency. International Journal for Research Publication and Seminar, 16(2), 163–181. https://doi.org/10.36676/jrps.v16.i2.262
18. Maria Thason, J. R., & Prasad, M. S. R. (2025). Bridging the Gap: Integrating Security into DevOps Practices for Enhanced Software Delivery and Risk Mitigation. International Journal of Research in Modern Engineering & Emerging Technology, 13(5), 1–22. https://doi.org/10.63345/ijrmeet.org.v13.i5.130501
19. Maria Thason, J. R., & Rastogi, D. D. (2025). Orchestrating Complex Release Pipelines in DevOps: Strategies for Managing Dependencies, Automation, and Continuous Delivery. Journal of Quantum Science and Technology, 2(2), 6–28. https://doi.org/10.63345/jqst.v2i2.292
20. Mokkapati, C., Goel, P., & Aggarwal, A. (2023). Scalable microservices architecture: Leadership approaches for high-performance retail systems. Darpan International Research Analysis, 11(1), 92.
21. Nadarajah, N., Ganipaneni, S., Chopra, P., Goel, O., & Goel, P. (2024). Achieving Operational Efficiency through Lean and Six Sigma Tools in Invoice Processing. Journal of Quantum Science and Technology, 1(3), 265–286.
22. Rajakumar Maria Thason, J., & Solanki, S. (2025). Automated Canary Deployments in Continuous Delivery: Balancing Speed and Reliability. International Journal of Enhanced Research in Science, Technology & Engineering, 14(4), 426–448.
23. Shekhar, E. S., Jain, A., & Goel, P. (2024). Building cloud-native architectures from scratch: Best practices and challenges. International Journal of Innovative Research in Technology, 9(6), 824–829.
24. Thason, J. R. M., & Singh, R. K. (2025). Measuring DevOps Success with the DORA Metrics: A Comprehensive Analysis of Key Performance Indicators and Their Impact on Software Delivery. International Journal of Creative Research Thoughts, 13(5), i341–i358. https://ijcrt.org/papers/IJCRT2505957.pdf
25. Vanam, L., Sachi, S., Gandhi, H., Shah, S. K., Kamballi, M., & Goel, P. (2025). Improving the Performance of Convolutional Neural Networks for Image Classification. In 2025 2nd International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS) (pp. 1–5). IEEE.
26. Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., and Liang, P. (2024). Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 12, 157–173. DOI: 10.1162/tacl_a_00638.
27. Maynez, J., Narayan, S., Bohnet, B., and McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1906–1919. DOI: 10.18653/v1/2020.acl-main.173.
28. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A., and Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), Article 248, 1–38. DOI: 10.1145/3571730.
29. Es, S., James, J., Espinosa Anke, L., and Schockaert, S. (2024). RAGAs: Automated evaluation of retrieval augmented generation. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, 150–158. DOI: 10.18653/v1/2024.eacl-demo.16.

OPEN ACCESS
@2026. This is an open access article distributed under the terms of the Creative Commons License [CC BY NC 4.0] and is available on www.wjcrt.org