International Journal of Multidisciplinary Allied Research Review and Practices (2455-1570)www.ijmarrp.in

BizLens: An Enterprise AI Decision Intelligence Platform with a Trustworthy AI Verification Layer

Tanmay Kshirsagar[1], Atharva Borate[1], Khushi Bhadangkar[1], Anushri Ghosh[1], Yuvaraj Nikam[2]
[1]Department of Computer Science & Engineering, MIT School of Computing, MIT ADT University, Pune, India
[2]Assistant Professor, Department of Computer Science & Engineering, MIT School of Computing, MIT ADT University, Pune, India
MIT School of Computing, MIT ADT University, Pune, India

Abstract
Modern businesses generate and store large volumes of information across heterogeneous formats such as Excel spreadsheets, CSV files, financial reports, sales reports, inventory records, and PDF documents. Extracting meaningful insights from these sources often requires significant manual effort for data cleaning, analysis, dashboard creation, trend identification, forecasting, and report interpretation. Existing Business Intelligence (BI) platforms and AI-powered analytics assistants have reduced the effort required for querying and visualization; however, AI-generated business insights can still contain incorrect calculations, hallucinated numerical values, unsupported claims, or contradictions between different data sources. These limitations become particularly significant when AI-generated outputs are used to support business decisions.
This paper presents BizLens, an Enterprise AI Decision Intelligence Platform that combines business analytics, Machine Learning (ML), Retrieval-Augmented Generation (RAG), multi-agent orchestration, and a Trustworthy AI Verification Layer. The platform accepts structured and unstructured business data, performs automated data profiling and analytics, generates dashboards, forecasts future trends, detects anomalies, and enables natural-language business question answering. The primary research contribution is its verification-centric architecture, which focuses on numeric provenance and cross-source reconciliation. Instead of directly presenting AI-generated responses, the system decomposes responses into atomic claims and independently validates important claims against deterministic analytics and retrieved source evidence. Numerical values are recalculated, business rules are checked, and information from multiple sources is compared to identify conflicts. Each generated insight can therefore be associated with verification status, supporting evidence, and a confidence score.
The proposed system uses a multi-agent architecture consisting of data ingestion, analytics, retrieval, insight generation, and verification agents. Hybrid retrieval combines keyword-based and semantic retrieval for business documents. An adversarial evaluation suite containing incorrect numbers, conflicting reports, incorrect calculations, and missing values is proposed to evaluate the reliability of the verification layer. The system aims to provide evidence-backed business intelligence while reducing the risk of unverified AI-generated information influencing business decisions.

Keywords: Business Intelligence, Decision Intelligence, Hallucination Detection, Retrieval-Augmented Generation, Trustworthy AI

I. INTRODUCTION

Businesses increasingly depend on data generated from sales systems, financial records, inventory management systems, customer interactions, human-resource reports, and operational activities. However, such information is often distributed across multiple file formats and sources, requiring analysts to clean spreadsheets, inspect reports, construct dashboards, identify trends, calculate key performance indicators, and interpret documents before reaching conclusions. Artificial Intelligence (AI) and Large Language Models (LLMs) have introduced new approaches to business analytics through natural-language querying, automated summaries, trend identification, and visualization. Retrieval-Augmented Generation (RAG) further enables language models to retrieve relevant organizational information before generating responses.

However, generating an answer does not guarantee that the answer is correct. AI-assisted business intelligence can produce hallucinated numerical values, incorrect calculations, insufficient provenance, and conflicts between multiple data sources. These limitations are particularly important when AI-generated outputs influence business

decisions. A fluent response may appear reasonable without indicating whether its numerical claims were independently calculated, supported by source evidence, or contradicted by another source. This creates an important distinction between generating an insight and trusting an insight.BizLens addresses this problem through a Trustworthy AI Verification Layer positioned between AI-generated insights and the final user-facing response. The proposed system decomposes generated responses into individual claims and independently verifies important claims using deterministic analytics and retrieved source evidence. Numerical values can be recalculated, business rules validated, and information from multiple sources compared to identify inconsistencies. Claims can then be associated with verification status, supporting evidence, and confidence information.

The research therefore focuses not on developing another general-purpose AI Business Intelligence assistant, but on a verification-gated business insight generation pipeline. The primary contribution is centered on numeric provenance and cross-source reconciliation, while multi-format data ingestion, business analytics, RAG, and multi-agent orchestration provide the supporting infrastructure.

II. LITERATURE REVIEW AND RELATED WORK

A. AI-Assisted Business Intelligence: AI-assisted Business Intelligence combines natural-language interfaces, automated visualization, Machine Learning, and generative AI. Platforms such as Power BI Copilot, ThoughtSpot, Amazon QuickSight, and Salesforce/Tableau already provide natural-language querying, automated insights, and conversational analytics. Therefore, BizLens cannot claim novelty from AI-powered BI alone. Many such platforms rely on structured or governed datasets, while smaller organizations may work with heterogeneous spreadsheets, CSV files, and PDF reports.

B. Retrieval-Augmented Generation: Retrieval-Augmented Generation (RAG) combines language-model generation with retrieved external information. In business applications, retrieval quality is important for finding numerical values, identifiers, dates, and business terms. BizLens uses hybrid retrieval with **BM25 and dense embeddings** to support exact and semantic matching, while associating responses with source evidence. However, retrieval alone cannot guarantee factual correctness, motivating its integration with the verification layer.

C. Hallucination and Trustworthy AI: Hallucination remains a significant challenge in Large Language Model applications. Research including RAGTruth, ReDeEP, HalluSearch, and MetaRAG has investigated hallucination detection and factual grounding.However, much of this research focuses on general question-answering and text generation, while Business Intelligence additionally requires numerical correctness, calculation consistency, source provenance, and cross-source agreement. BizLens therefore adopts an output-level verification approach in which generated claims are evaluated against source evidence and deterministic computations.

D. Existing Open-Source AI-BI Systems: Open-source systems such as DB-GPT, WrenAI, OpenChatBI, and querycraft-ai demonstrate established capabilities including natural-language querying, RAG, agentic workflows, NL-to-SQL, and dashboards. Therefore, basic AI-BI functionality is not the primary contribution of BizLens. Instead, BizLens focuses on numerical provenance, independent recalculation, and cross-source conflict detection to improve the reliability of AI-generated business insights.

III. PROBLEM STATEMENT AND RESEARCH GAP

A. Problem Statement

Business information is commonly distributed across Excel spreadsheets, CSV files, PDF reports, and other business documents. Analysts must process these heterogeneous sources to identify trends, calculate Key Performance Indicators (KPIs), build dashboards, forecast performance, and interpret reports. Although AI-powered Business Intelligence tools can automate several of these activities, AI-generated answers may contain hallucinated numbers,

incorrect calculations, unsupported statements, or contradictions between different sources. Existing systems may therefore generate plausible responses without independently validating important numerical or factual claims.The BizLens problem can therefore be formulated as follows:How can an AI-powered Business Intelligence system generate useful business insights from heterogeneous structured and unstructured data while independently verifying the factual and numerical correctness of those insights before presenting them to users?BizLens addresses this problem through multi-format data processing, deterministic analytics, Retrieval-Augmented Generation (RAG), and a verification-gated output pipeline.

B. Research Gap

Research indicates that several components of AI-assisted Business Intelligence are already relatively mature, including NL-to-SQL, natural-language-to-chart generation, document RAG, dashboard generation, and schema-constrained query generation.The more significant gaps identified are:

1. Sentence-level numerical provenance: Generated numerical claims are not consistently associated with the exact computed value or source evidence supporting them.
2. Cross-source reconciliation: Existing systems are not primarily designed to detect and explicitly flag conflicts when different business sources report different values for the same metric.
3. Heterogeneous raw-data ingestion: Many systems assume a governed semantic layer or warehouse rather than collections of raw Excel, CSV, and PDF files.
4. Business-specific hallucination evaluation: Existing hallucination benchmarks often focus on prose-based question answering rather than numerical consistency between generated responses and structured business datasets.

The proposed BizLens contribution addresses these gaps through a verification-gated business insight pipeline designed to improve the reliability and transparency of AI-generated business insights. The primary research focus is on numeric provenance and cross-source reconciliation, while deterministic analytics, multi-format ingestion, RAG, and multi-agent orchestration provide the supporting infrastructure.

IV. OBJECTIVES AND CONTRIBUTIONS

The primary objective of BizLens is to develop an **Enterprise AI Decision Intelligence Platform** capable of analyzing heterogeneous business data while improving the trustworthiness and transparency of AI-generated business insights. The objectives are:

1. To support ingestion of Excel, CSV, and PDF business data.
2. To profile data quality and identify missing values, duplicate records, inconsistent data types, outliers, and other data-quality issues.
3. To perform exploratory data analysis, KPI generation, trend analysis, anomaly detection, and forecasting.
4. To provide interactive business dashboards for presenting analytical results.
5. To support natural-language business question answering using Retrieval-Augmented Generation (RAG).
6. To provide citations and source evidence for document-based answers.
7. To independently verify important AI-generated numerical claims against deterministic computations and available source evidence.
8. To validate generated claims against predefined business rules.
9. To detect and flag conflicting information across uploaded business sources.
10. To provide claim-level verification status, supporting evidence, and confidence information.
11. To evaluate the verification layer using intentionally corrupted, incorrect, and conflicting test data.

The principal contribution of BizLens is a **verification-gated business-insight generation pipeline** in which important numerical claims generated by an LLM are independently checked against deterministic computations and source evidence. The system also identifies conflicts across structured and unstructured sources and communicates claim-level verification status and supporting evidence. Thus, **numeric provenance, independent verification, and cross-source reconciliation** form the primary research contribution, while analytics, RAG, multi-format ingestion, and multi-agent orchestration provide the supporting platform capabilities.

V. PROPOSED BIZLENS FRAMEWORK

BizLens BizLens is designed as an Enterprise AI Decision Intelligence Platform that combines multi-format data processing, deterministic business analytics, Machine Learning (ML), Retrieval-Augmented Generation (RAG), multi-agent orchestration, and a Trustworthy AI Verification Layer. The architecture separates insight generation from insight verification so that important AI-generated claims can be independently evaluated using analytical results and source evidence.

A. Overall Framework

The proposed BizLens workflow is:

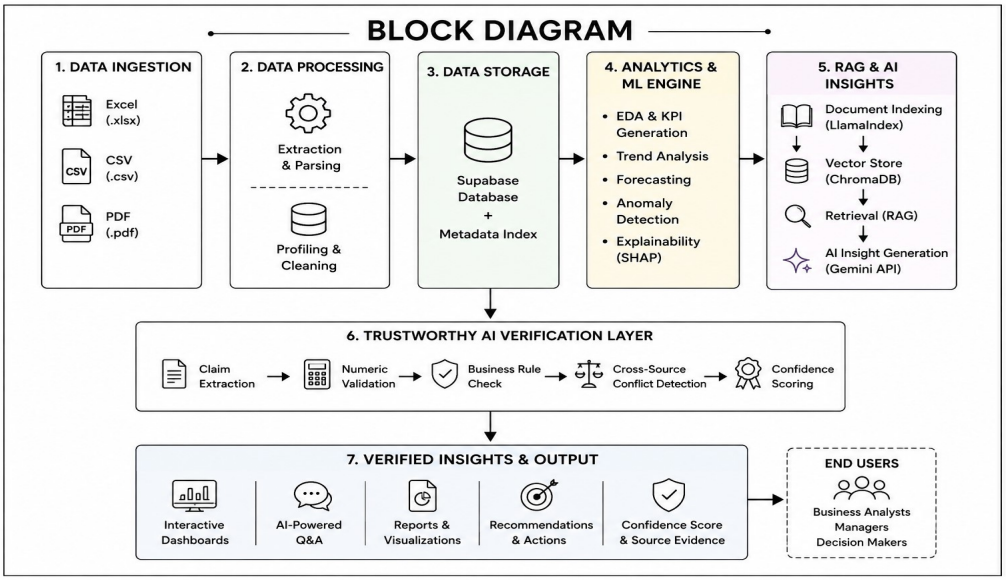

Fig I. System Architecture

The platform accepts structured and unstructured business information, including Excel, CSV, PDF reports, financial reports, sales reports, inventory reports, and other business documents. Uploaded data is processed and transformed into representations suitable for analytical processing and document retrieval.The data profiling stage identifies missing values, duplicate records, inconsistent data types, outliers, and other data-quality issues. The deterministic analytics layer then performs exploratory data analysis, Key Performance Indicator (KPI) generation, trend analysis, forecasting, statistical analysis, and anomaly detection.

For unstructured documents, BizLens uses Retrieval-Augmented Generation with hybrid retrieval combining BM25-based lexical retrieval and dense semantic embeddings. This supports both exact matching of business-specific terms, identifiers, and numerical information and semantic retrieval of relevant content. Retrieved evidence is supplied to the generation layer and associated with relevant source citations.The Insight Generation Agent combines analytical results and retrieved evidence to produce natural-language business insights, summaries, and answers. Generated output is not treated as automatically trustworthy. Important claims are extracted and passed to the verification stage before verification information is communicated to the user.

B. Multi-Agent Architecture: BizLens uses specialized agents to separate major system responsibilities:

1. **Data Ingestion Agent:** Processes uploaded business files, identifies relevant structures and schemas, performs preprocessing where applicable, and produces representations suitable for downstream processing.
2. **Analytics Agent:** Performs deterministic business analysis, including exploratory data analysis, KPI generation, trend analysis, forecasting, and anomaly detection. Its outputs also provide independently computed values for numerical verification.
3. **Retrieval Agent:** Processes unstructured business documents and performs hybrid retrieval using lexical and dense retrieval methods. It provides relevant information and source evidence for business queries.
4. **Insight Generation Agent:** Combines analytical results and retrieved evidence to generate natural-language business insights and answers.
5. **5Verification Agent:** Evaluates important generated claims by verifying AI-generated values, recalculating numerical results, checking predefined business rules, detecting conflicts across sources, and assigning verification status and confidence information.

C. Trustworthy AI Verification Layer: The Trustworthy AI Verification Layer represents the central research contribution of BizLens. Instead of evaluating an AI-generated response as a single output, important claims are evaluated individually against independently available evidence and deterministic computations.

6. **Atomic Claim Verification**: Generated responses are decomposed into individual factual or numerical claims. For example:Revenue increased by 12%. The North region performed best. Profit increased by 20%.
7. **Independent Numerical Verification**: Numerical claims such as revenue growth, profit margins, and percentage changes are independently recalculated by the deterministic analytics layer.
8. **3. Business Rule Validation:** Generated and calculated information is checked against predefined business rules. Violations are flagged as potential data inconsistencies, calculation errors, or unsupported claims.
9. **Cross-Source Conflict Detection:** BizLens compares information across business sources and flags conflicting values rather than arbitrarily selecting one.
10. **Verification Status and Confidence:** Important claims are associated with verification status, supporting evidence, and confidence information based on numerical validation and source agreement.

D. End-to-End Processing: BizLens separates **insight generation** from **verification**. Analytical results and retrieved evidence are provided to the Insight Generation Agent, while generated claims are independently evaluated by the Verification Agent using calculations, business rules, and source comparison.The final response presents relevant evidence and verification information, while unsupported or conflicting claims are flagged. This positions the LLM as an **insight-generation component rather than the final authority**.

VI. IMPLEMENTATION AND TECHNOLOGY STACK

The BizLens implementation uses a modular technology stack supporting data ingestion, business analytics, retrieval, AI-based insight generation, verification, and visualization. The frontend uses **React.js**, **TypeScript**, **Tailwind CSS**, **Plotly**, and **Chart.js** for interactive interfaces and data visualization, while **Python** and **FastAPI**

provide the backend and API layer. **PostgreSQL** and **Supabase** support data storage and backend services, with **Supabase Authentication** providing user authentication and access management. The AI and retrieval layer uses the **Gemini API** for Large Language Model capabilities, **LlamaIndex** for document processing and RAG workflows, **BM25** and **dense embeddings** for hybrid retrieval, and **ChromaDB** for vector storage. **LangGraph** orchestrates the specialized agents and conditional workflows. **Pandas**, **NumPy**, and **Scikit-learn** support data processing and Machine Learning, while **SHAP** provides model interpretability. For deployment, **Vercel** is used for the frontend and **Render** for the backend.

VII. EXPERIMENTAL METHODOLOGY AND EVALUATION

A research-oriented evaluation is necessary to assess whether BizLens provides measurable improvements in the reliability of AI-generated business insights rather than functioning only as a software demonstration. The evaluation focuses primarily on the **Trustworthy AI Verification Layer**, with retrieval and forecasting evaluated where applicable.An **Adversarial Test Suite** is constructed using deliberately modified datasets and business reports. The test cases include:

1. **Incorrect numerical values** – intentionally modified values to test numerical verification.
2. **Conflicting values between reports** – inconsistent values for the same metric across different sources.
3. **Incorrect calculations** – deliberately incorrect derived values to evaluate recalculation-based verification.
4. **Missing values** – incomplete records to examine handling of insufficient information.
5. **Inconsistent business rules** – violations of predefined business constraints.

The system is evaluated based on its ability to identify, flag, and correctly classify these inconsistencies.

A. Golden Evaluation Set: A manually verified **Golden Evaluation Set** is constructed containing business questions and expected answers. It includes questions that can be deterministically calculated from structured datasets and questions directly supported by uploaded business documents. The set provides a consistent reference for evaluating generated and verified claims.

B. Verification Accuracy: Verification accuracy measures the proportion of evaluated claims for which the system correctly determines the verification outcome:Verification Accuracy=Correctly Verified Claims/ Total Evaluated Claims×100. This metric provides a direct measure of the effectiveness of the proposed verification layer.

C. Conflict Detection: Cross-source conflict detection is evaluated using **precision, recall, and F1-score** over intentionally introduced conflicts. These metrics measure the system's ability to identify genuine conflicts while minimizing false detections.

D. Retrieval Evaluation: Because retrieved evidence affects the reliability of generated responses, the retrieval component is evaluated using precision, recall, and relevant-document retrieval rate. Evaluation focuses on whether the hybrid retrieval mechanism retrieves the documents or passages containing information required to answer the evaluated business questions.

VIII. RESULTS AND ANALYSIS

The performance of BizLens was evaluated using the experimental methodology described in the previous section. The evaluation focused on retrieval quality, cross-source conflict detection, verification accuracy, hallucination suppression, and system response time.

Metric	BizLens
Retrieval Precision@5	0.884
Retrieval Recall@5	0.80
Conflict Detection F1	0.856
Verification Accuracy	91.0%
Hallucination Suppression	90.8%
Average Response Time	2.38 s

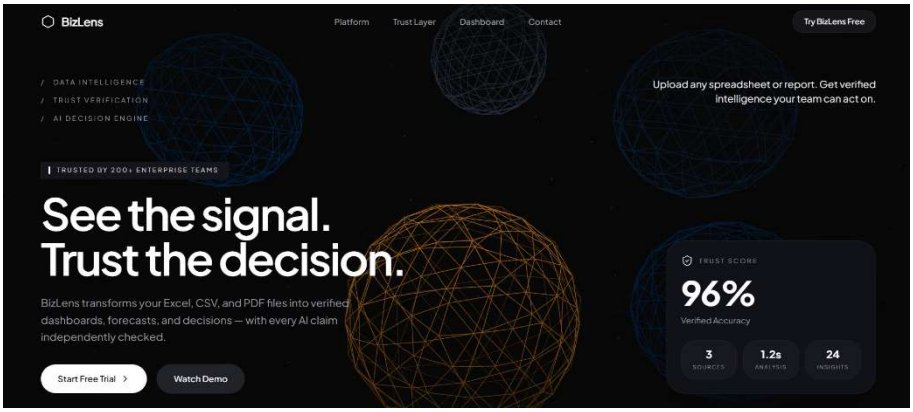

Fig II. Home Page

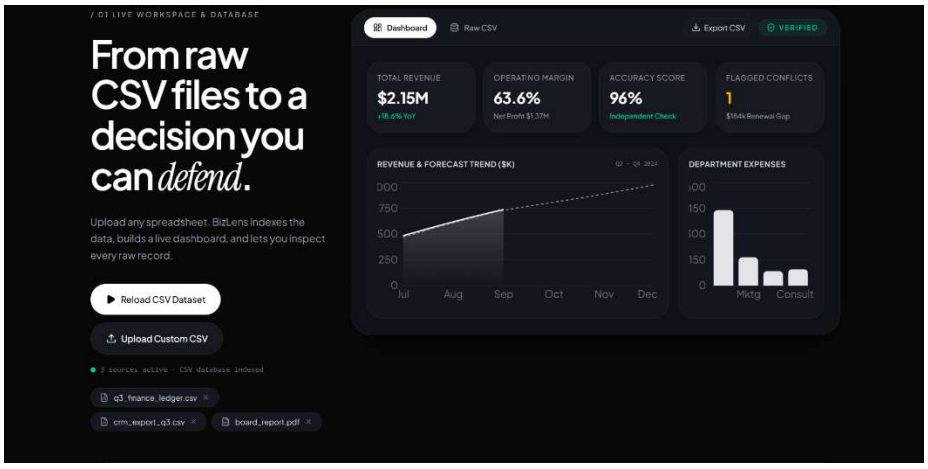

Fig III. Framework

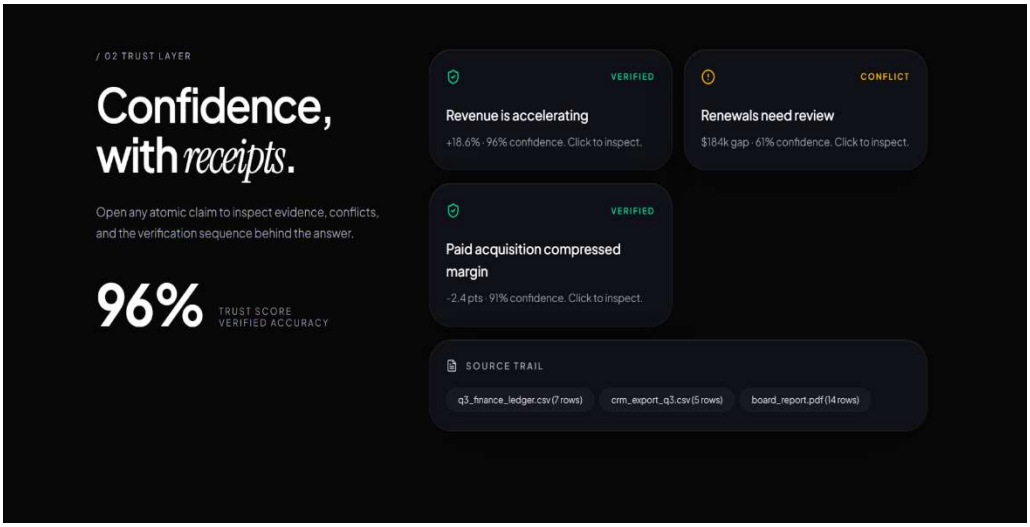

Fig IV. Confidence Score

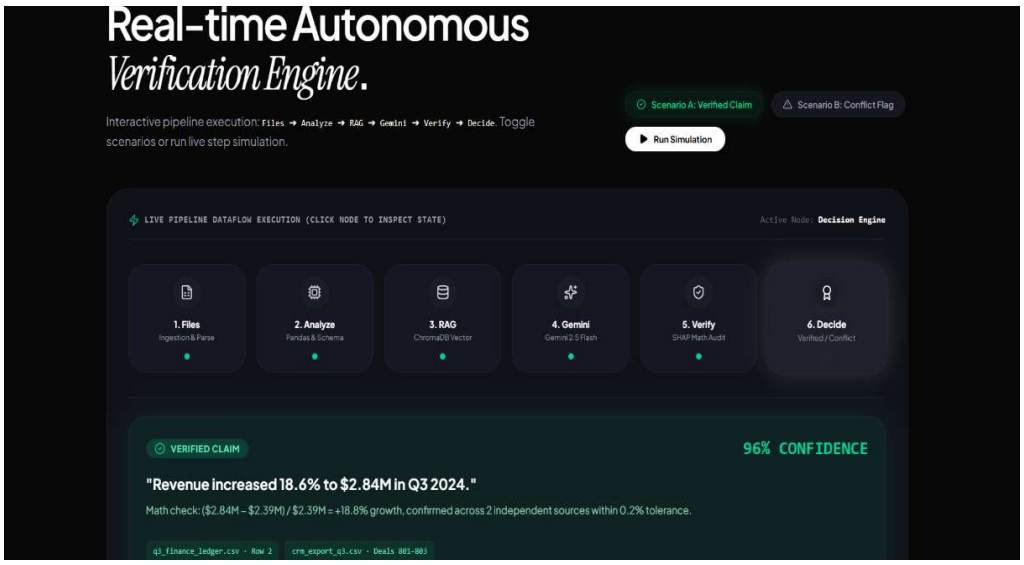

Fig V. Verification Layer

IX. DISCUSSION

The proposed architecture changes the role of generative AI within Business Intelligence by treating the LLM as an insight-generation component rather than the final authority. Generated claims can be independently evaluated using deterministic calculations, retrieved evidence, business rules, and cross-source comparison.This approach is particularly relevant for numerical business information, where a fluent AI-generated response may still contain incorrect values or calculations. By separating generation from verification, BizLens is designed to expose unsupported claims and conflicts rather than silently presenting potentially unreliable information.

X. LIMITATIONS

The proposed system has several limitations. Its reliability remains dependent on the quality and completeness of the input data and retrieved evidence; verification cannot guarantee that an original business source is itself correct. The current implementation is focused on file-based sources such as Excel, CSV, and PDF rather than real-time enterprise data streams. The ChromaDB-based retrieval infrastructure is primarily intended for development and student-project scale rather than large-scale multi-tenant deployment. In addition, confidential business data introduces security and privacy requirements that must be addressed through appropriate access control and data-protection mechanisms. Finally, confidence information should not be interpreted as an absolute probability of correctness unless it is empirically calibrated against a sufficiently large evaluation dataset.

XI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

BizLens proposes an Enterprise AI Decision Intelligence Platform that combines heterogeneous business-data ingestion, deterministic analytics, Machine Learning, Retrieval-Augmented Generation, multi-agent orchestration, and a Trustworthy AI Verification Layer.

The primary motivation is that generating an AI-based business insight is not sufficient when the insight may influence operational or strategic decisions. AI-generated responses can contain incorrect calculations, unsupported claims, or information that conflicts with other business sources. BizLens addresses this challenge by separating **insight generation from insight verification**.The verification layer decomposes AI responses into atomic claims, independently validates numerical values, checks business rules, identifies conflicts between structured and unstructured sources, and provides claim-level verification status, evidence, and confidence information. Thus, BizLens focuses on moving from conventional AI-assisted Business Intelligence toward **verification-aware decision intelligence**.The central research contribution is therefore positioned around **numeric provenance, cross-source reconciliation, and verification-gated insight generation**, rather than the general concept of an AI Business Intelligence assistant. Final effectiveness will be established through the experimental evaluation described in the methodology using verification accuracy, retrieval performance, conflict-detection performance, and other applicable metrics.

B. Future Scope

Future development can extend BizLens through knowledge graphs, real-time data streaming, stronger role-based access control, audit logging, semantic caching, model routing, and specialized model fine-tuning. These extensions could improve contextual understanding, scalability, security, and adaptability for larger enterprise environments.Multi-tenant production deployment can also be explored to support multiple organizations, users, and isolated business datasets. Additional research can investigate more advanced verification techniques, including richer relationships between business entities and metrics and improved methods for calibrating confidence information.Future research should also develop larger evaluation datasets containing structured and unstructured business sources with deliberately introduced numerical inconsistencies. Such datasets would enable more rigorous benchmarking of business-specific hallucination detection, numerical verification, and cross-source reconciliation. This would help establish standardized evaluation practices for trustworthy AI systems operating on business and tabular data.Overall, the future direction of BizLens is to evolve from a project-scale decision intelligence platform into a more scalable and rigorously evaluated trustworthy AI framework for enterprise business analytics.

REFERENCES

[1] C. Niu, Y. Wu, J. Zhu, S. Xu, K. Shum, R. Zhong, J. Song, and T. Zhang, "RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models," *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pp. 10862–10878, 2024, doi: 10.18653/v1/2024.acl-long.585.

[2] Z. Sun, X. Zang, K. Zheng, J. Xu, X. Zhang, W. Yu, Y. Song, and H. Li, "ReDeEP: Detecting Hallucination in Retrieval-Augmented Generation via Mechanistic Interpretability," *International Conference on Learning Representations*, 2025.

[3] M. Abdallah and S. El-Beltagy, "HalluSearch at SemEval-2025 Task 3: A Search-Enhanced RAG Pipeline for Hallucination Detection," *Proceedings of the 19th International Workshop on Semantic Evaluation*, pp. 1436–1441, 2025.

[4] C. Sok, D. Luz, and Y. Haddam, "MetaRAG: Metamorphic Testing for Hallucination Detection in RAG Systems," arXiv preprint arXiv:2509.09360, 2025.

[5] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Transactions on Information Systems*, vol. 43, no. 2, pp. 1–55, 2025, doi: 10.1145/3703155.

[6] A. Jacovi et al., "The FACTS Grounding Leaderboard: Benchmarking LLMs' Ability to Ground Responses to Long-Form Input," 2025.

[7] S. Xue et al., "DB-GPT: Empowering Database Interactions with Private Large Language Models," arXiv preprint arXiv:2312.17449, 2023.

[8] Canner, "WrenAI: Open-Source Generative Business Intelligence Platform," GitHub repository.

[9] "OpenChatBI: Conversational Business Intelligence System," project repository/documentation. [Verify original publication/repository metadata before final submission.]

[10] "querycraft-ai: Schema-Constrained Multi-Agent Query Generation and Validation," project repository. [Verify original metadata before final submission.]

[11] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[12] A. Asai et al., "Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection," *International Conference on Learning Representations*, 2024.

[13] T. Gao et al., "Precise Zero-Shot Dense Retrieval without Relevance Labels," *ACL*, 2021.
(Useful if you discuss dense retrieval/embeddings.)