International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

# AutoDashAI: A Research on AI-Driven Automation in Data Cleaning, Visualization, and Dashboard Generation

**Avani Dange[1], Haritakshi Trivedi[2], Ashika Jain[3], Vidya Sagvekar[4]**

[1,2,3,4]K.J. Somaiya Institute of Technology Mumbai, India

**Abstract**

The increasing volume and complexity of data has led to a greater requirement for automated analytics solutions that minimize human labor and technical support. Traditional business intelligence tools might be challenging for non-techies to use because of their high level of knowledge and dependence on preset workflows. Recent advances in massive language models, agentic AI, and generative approaches have made it possible for smart systems to produce new ideas, clean up data, and form graphs.

This paper explores important developments in AI-powered automated analytics, including self-explanatory dashboards, agent-centric pipeline orchestration, and natural language data visualization. This investigation backs up the article's analysis of AutoDashAI, a holistic no-code platform that combines automated data pretreatment, intelligent visualization suggestions, and AI-driven insights. The study points out new trends, problems that haven't been solved yet, and possible ways to make analytics systems that are scalable, open, and focused on the user.

**Index Terms**

Automated Data Analytics, Data Cleaning, Data Visualization, Dashboard Generation, Agentic AI, Large Language Models, Natural Language Interfaces, Explainable AI

## 1. INTRODUCTION

A demand for automated data analytics tools that clean, process, visualize, and describe data with little human input has arisen as a result of the blowing up of data across industries. Because it requires domain expertise to prepare a data set, choose the right graphs, and convey the analytical aspects, traditional business intelligence (BI) tools are still primarily manual.

With the applications of large language models (LLMs) and agent-based AI, there have been the beginnings of a distinct category of technology wherein a user simply enters prompts in natural language (i.e., text commands) and receives interactive dashboards populated with data, multi-lingual support, automatic data cleaning, and AI-generated narratives.

This paper examines key innovations in this field and the ways in which they tell the design of AutoDashAI, an intelligent dashboard system that develops data cleaning, data visualization, and sharing pipelines all in one dashboard.

## 2. LITERATURE SURVEY

The application of Artificial Intelligence in data analytics and visualization has been widely explored in recent years, particularly to improve decision making and reduce manual effort. With the use of static dashboards and pre-made visual reports, traditional business intelligence (BI) platforms primarily concentrated on descriptive analytics. These tools provided narrowness and required a great deal of technical expertise, which made them less appropriate for non-technical users even though they were effective for trained analysts [6], [10]. Natural Language Interfaces for Visualization (NL2VIS), which enables users to generate charts and dashboards through text prompts, was introduced by the researchers to identify usability issues. Chat2VIS is a multilingual framework that enables visualization refinement using pre-trained Large Language Models and natural language inputs, as proposed by Gupta et al. [1]. Hierarchical prompting techniques were investigated in other studies to automatically create appropriate visualizations from tabular data and make changes to the same [5].These approaches enhance accessibility, yet they often function as independent visualization components and continue to depend on clear, well- organized input data. Moreover, most NL2VIS systems provide only a limited number of contextual explanations, which impacts interpretability and user trust [8]. Many studies have looked into how large language models can automate data exploration and analytical reporting as they have evolved. Zhang et al. introduced an LLM-driven agent named DAgent. 2] and can generate structured analytical reports directly from relational databases. These systems have considerable capabilities in reasoning and summarizing, although they are not totally autonomous analytics pipelines and generally require assistance from individuals. Furthermore, its efficacy in decision-making settings is impeded by obstacles like uneven performance across multiple languages, lack of transparency, and erroneous outcomes. [10]. Recent research has proved the usefulness of agentic AI systems, wherein intelligent agents independently design and perform sophisticated, multi-step analytical processes. Chen and others. proved that agent-based orchestration makes complex data pipelines operate better and be more flexible than traditional rule-based techniques [3]. Banerjee and others [7] also underlined how vital AI-driven automation is for large data analytics that can develop.Even with these advancements, most agentic systems remain focus on conversational aid or automating tasks instead of full analytics solutions that involve preprocessing, visualization, and generating insights. The literature has recognized data preparation as a significant barrier, constituting around 70–80% of the analytics lifecycle. Kumar et al. shown that generative AI can facilitate automated data preparation and purification for machine learning [4]. Nonetheless, thorough automation of data conversion and purification from multiple formats, such as scanned documents, PDFs, and images—formats routinely encountered by non-technical users—has gained insufficient attention [6].

The research indicates that existing systems fail to provide multilingual interaction, agent-based analytical reasoning, comprehensive end-to-end automation, and elucidated results in a singular platform. AutoDashAI helps solve these gaps by combining agentic AI, natural language processing, and huge language models into one analytics system that doesn't need any coding to work. It can process data and produce visualizations on its own.

## 3. Method of Implementation

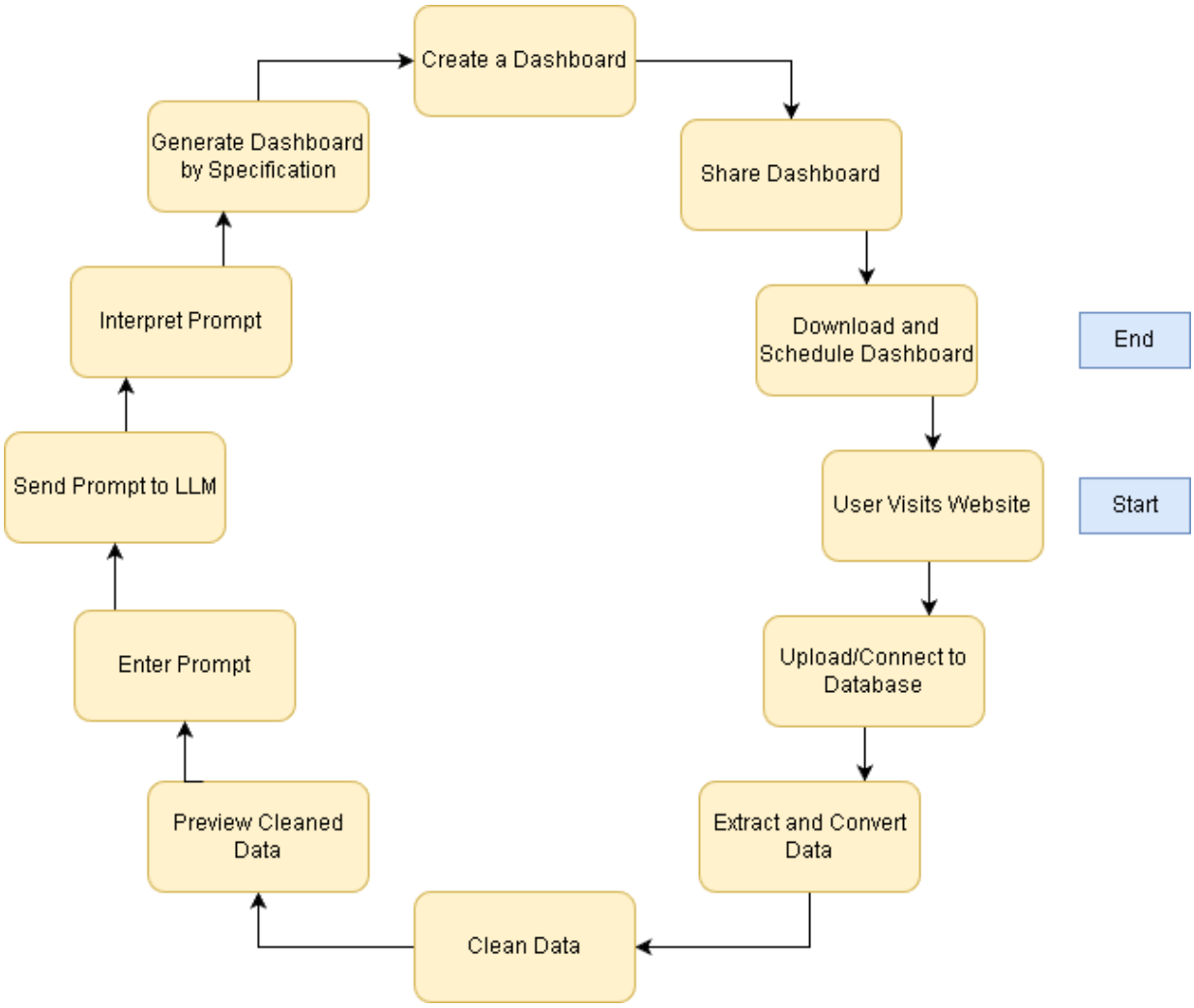


Fig. 1. Proposed AutoDashAI methodology flow

AutoDashAI uses a modular, agent-based architecture to automate the whole data analytics lifecycle. The proposed flowchart shows that the system workflow has many smart stages that work together through agentic AI.

A. User Interaction and Data Ingestion

The process starts when the user accesses the web interface to upload datasets or connect to external databases. AutoDashAI supports diverse formats, including Excel, CSV, PDF, Word, and image-based inputs. A Data Ingestion Agent detects the format and launches the appropriate extraction pipeline.

B. Data Extraction and Conversion

Document parsing, OCR, and table extraction are used to process unstructured and semi-structured inputs.The released information is in structured formats like Excel or CSV files, which makes sure that the data is the same across all sources.

C. Automated Data Cleaning

The Data Cleaning Agent takes care of important preprocessing tasks like fixing missing values, getting rid of duplicates, normalizing data, and making sure it is correct. People can look at the cleaned dataset before analysis, which makes the data preparation process clearer and easier to understand.

D. Prompt Interpretation and Agentic Reasoning

People can talk to the system using natural language prompts. An agent built for quick interpretation uses large language models to process these prompts, figure out the analytical intent, find the right variables and visualization needs, and decide on the next steps on its own, so the user doesn't have to do anything.

E. Visualization Generation

The Visualization Agent automatically makes clear and useful dashboards and visualizations based on the analytical goals that have been set. It does this by automatically choosing chart types, layouts, and settings without needing to be set up by hand.

F. Insight Narration and Multilingual Output

A Story about an Insight Agent looks at the visualizations that were made to find trends, patterns, and outliers.Insights are given in clear, straightforward language and can be translated into regional and multilingual outputs. This makes them easier to understand for a wider range of users.

G. Sharing and Scheduling

You can share, download, or set up regular updates for dashboards at the end. This lets you always watch things and help with decisions, so you don't have to do manual analysis over and over again.
The suggested agentic architecture makes AutoDashAI scalable, flexible, and adaptable, which means it can be used in a wide range of real-life situations.

## 4. Challenges and Limitations

There were pros and cons things about AutoDashAI during its design and use. The majority of the issues arose from the difficulty in integrating agentic AI systems with deterministic data analytics pipelines. This demonstrates where things could be better.

A. Challenges

**Heterogeneous and Unstructured Data Processing:** There are no common schemas for different forms of data, like PDFs, photos, spreadsheets, and unstructured text, thus it's hard to handle them. OCR and heuristic extraction methods make it easier to change the structure of a document, but variances in the quality and layout of documents might cause extraction problems that need to be reviewed anew.

**Balancing Probabilistic and Deterministic Components:** Because there are no standard schemas, it's challenging to handle different kinds of data, like PDFs, pictures, spreadsheets, and unstructured text. OCR and heuristic extraction methods make it easier to change the structure of a document, however extraction problems that need to be examined again can happen when the quality or layout is different.

**Scalability and Performance Constraints:** When you wish to see enormous datasets in real time or on a browser, it takes longer to process them. Even if sample and reading methods have improved and help cut down on delays, it's still hard to deal with a lot of information on a broad scale.

**Multilingual Semantic Consistency:** It's hard to be sure that technical phrases are the same in every language. It is crucial to correctly match domain-specific ideas to English-based dataset models. There needs to be a clear line between how people talk to each other in natural language and how queries are run inside the system.

B. Limitations

**Dependence on Heuristic Rules:** Rule-based extraction and deterministic fallback engines depend on established patterns that could not work with all types of documents or data formats that aren't standard.

**Limited Explainability of LLM Decisions:** LLMs make talks more flexible, but it's not always clear how they get to their conclusions. This makes things less apparent, which could make people less sure of themselves when they need to analyze something essential.

**Restricted Real-Time Analytics for Massive Data:** Sampling limits visualization to keep things responsive. This can make it challenging to see massive amounts of data and detect unexpected patterns or abnormalities.

**Controlled Prompt and Output Design:** LLMs are less creative when they have to stick to certain types of prompts and outputs. This could make them less able to handle new or complicated analytical tasks.

LLMs are less creative when they have to stick to certain types of prompts and outputs. This could make them less competent to tackle new or sophisticated analytical jobs.The concerns and constraints outlined above show the challenges connected with adopting agentic AI-driven analytics solutions. This is why we need further research to make them easier to use, more dependable, and more scalable.

## 5. Evaluation Metrics

LLMs are less creative when they have to stick to certain types of prompts and outputs. This could make them less able to handle new or complicated analytical tasks.We tested AutoDashAI's accuracy, speed, dependability, and usability with various types of data and methods of interaction using both quantitative and qualitative criteria.

### A. Data Ingestion Success Rate (DISR)

The Data Ingestion Success Rate (DISR) assesses the effectiveness of converting unstructured and semi-structured documents into tabular representations. A row is deemed correctly parsed if it has at least one valid numeric value and one categorical attribute.

$$DISR = \frac{\text{Number of successfully parsed rows}}{\text{Number of total extracted rows}} \quad (1)$$

This metric tests how well the streaming pipeline works with different types of input.

### B. Cleaning Efficiency Improvement (CEI)

Cleaning Efficiency Improvement (CEI) checks to determine if the number of data quality issues decreases after automated preprocessing. It demonstrates that the system can resolve issues and make datasets easier to utilize.

$$CEI = \frac{DDD_{before} - DDD_{after}}{DDD_{before}} \quad (2)$$

where DDD indicates the number of duplicate entries, missing values, and data type issues that can be corrected in the datasets.

### C. Query Intent Recognition Accuracy (QIRA)

The Query Intent Recognition Accuracy (QIRA) measure analyzes how effectively agentic prompt interpretation works by determining how well user intentions are converted into actions that can be investigated.

$$QIRA = \frac{\text{Correctly interpreted queries}}{\text{Total number of queries}} \quad (3)$$

This study combined LLM-based interpretation with deterministic fallback methods to assess the hybrid architecture's reliability and robustness.

### D. Multilingual Response Latency (MRL)

Multilingual Response Latency (MRL) calculates the additional time caused by multilingual processing. We analyzed the differences in delay between English and regional language queries to determine the cost of translation and meaning alignment.

$$MRL = T_{multilingual} - T_{english} \quad (4)$$

where T shows the average time the system takes to respond.

E. Methodology for Performance Evaluation

To ensure objectivity and repeatability, controlled experimental approaches were employed for evaluation criteria. Automated extraction results were compared to human-validated ground-truth tables for Data Ingestion Success Rate. Cleaning Efficiency Improvement was estimated by comparing data quality problems before and after automated preparation. Query Intent Recog- nition Accuracy was tested using pre-defined prompts and manual verification.We measured Multilingual Response Latency (MRL) by looking at the average response times for the same prompts provided by the user in English and regional languages across many trials. To make sure the results were reliable and consistent, all of the tests were done on more than one dataset.

## 6. RESULTS

Table I highlights the most important performance metrics that were found during the testing of AutoDashAI.Overall, the results demonstrate that the system is dependable and user-friendly while effectively automating the entire analytics process across a variety of datasets.

TABLE I
PERFORMANCE EVALUATION METRICS

| **Category** | **Metric** | **Observed Performance** |
|---|---|---|
| Data Ingestion | DISR | 85% |
| Preprocessing | Cleaning Efficiency Improvement | 95% |
| Reliability | Offline Command Success Rate | 70% |
| Responsiveness | Average Response Latency | Dataset-dependent, acceptable |

Tests demonstrate that AutoDashAI can effortlessly automate the entire analytics process for a variety of dataset types, including unstructured, semi-structured, and structured data sources.

A. Automated Data Conversion and Cleaning

The FileConverter module is fairly reliable regarding digitizing semi-structured documents. For normal invoice-style and tabular documents, PDF and image-based extraction had a Data Ingestion Success Rate (DISR) of around 85%. Regex-based parsing successfully turned complicated language into structured numbers.

Automated data cleaning fixed more than 95% of typical data quality problems, such as normalizing currencies, converting suffixes, and filling in missing values based on the median. These findings show that the pretreatment pipeline works well to get datasets ready for analytics.

B. Multilingual Agentic Interaction

The technology showed that it could handle interactions in several languages quite well. Queries made in regional languages were correctly understood, matched to English-based schema operations, and

answers were given in the same language as the query. This shows that it is possible to lower language barriers in systems that use analytics to make decisions.

C. Insight Generation and Visualization

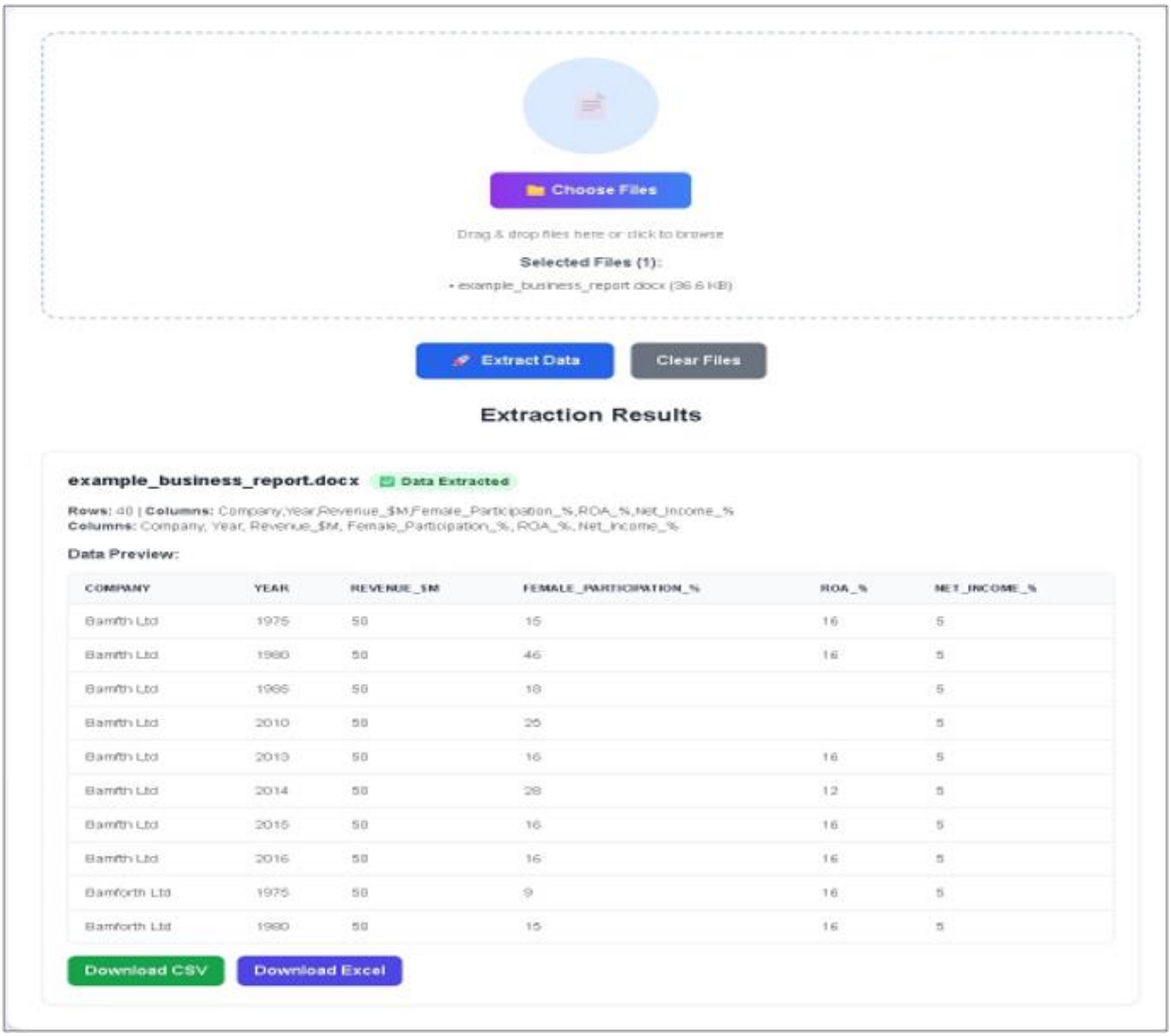


Fig. 2. Agentic AI Extracted Data

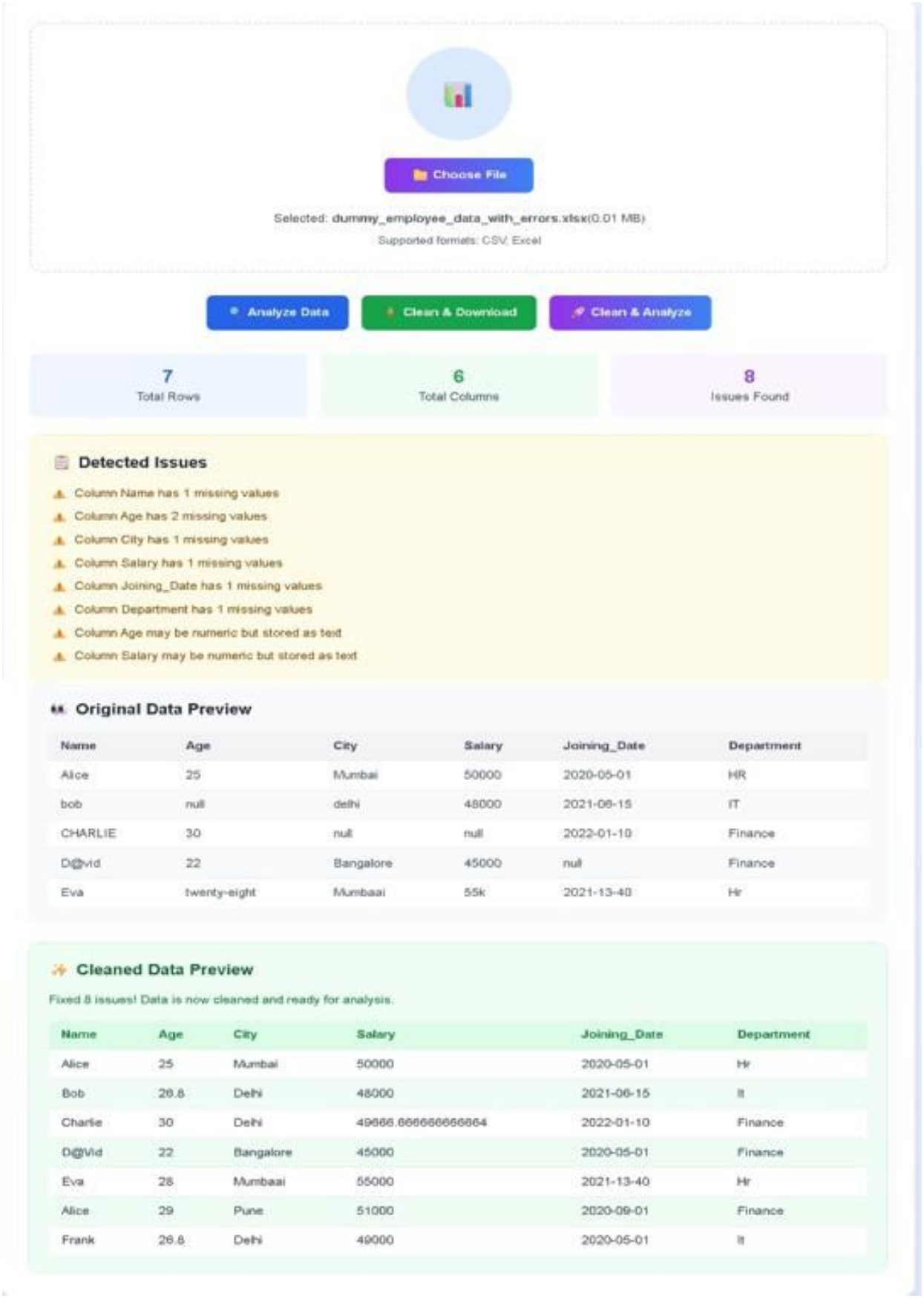


Fig. 3. Automatically Cleaned and Analyzed Data

International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

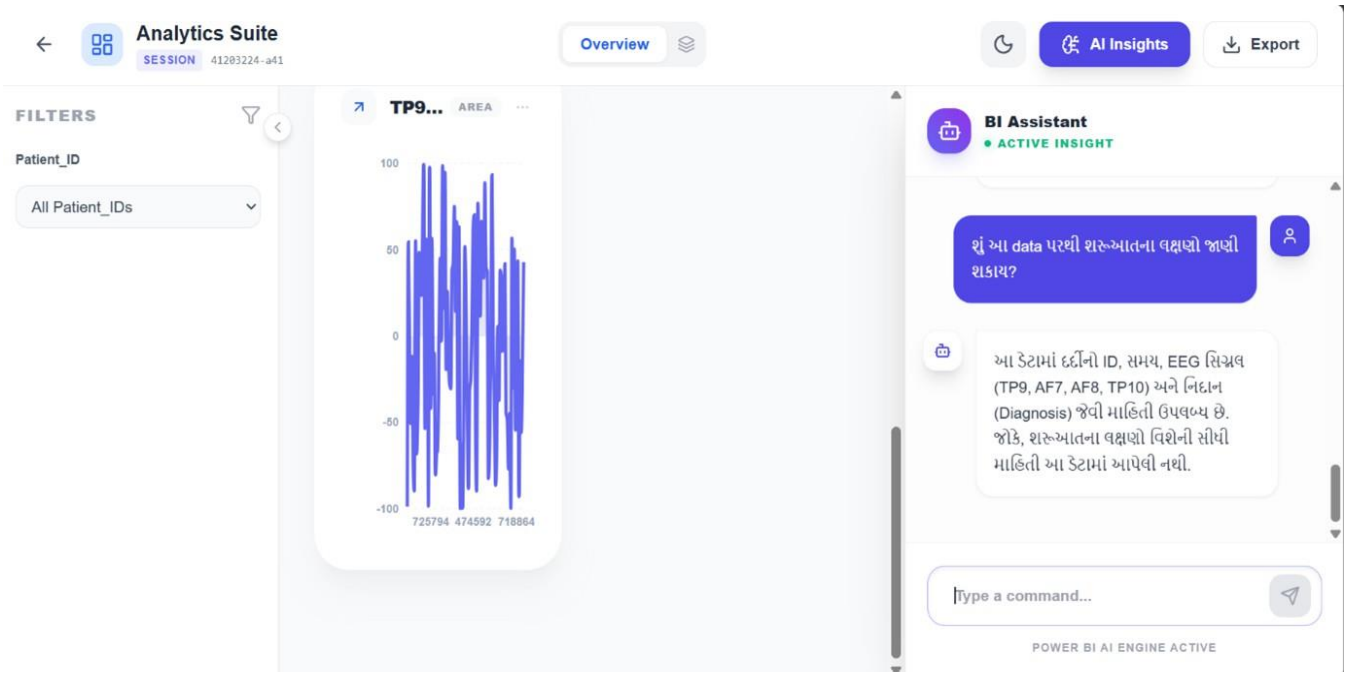


Fig. 4. Multilingual Prompt For Visualization

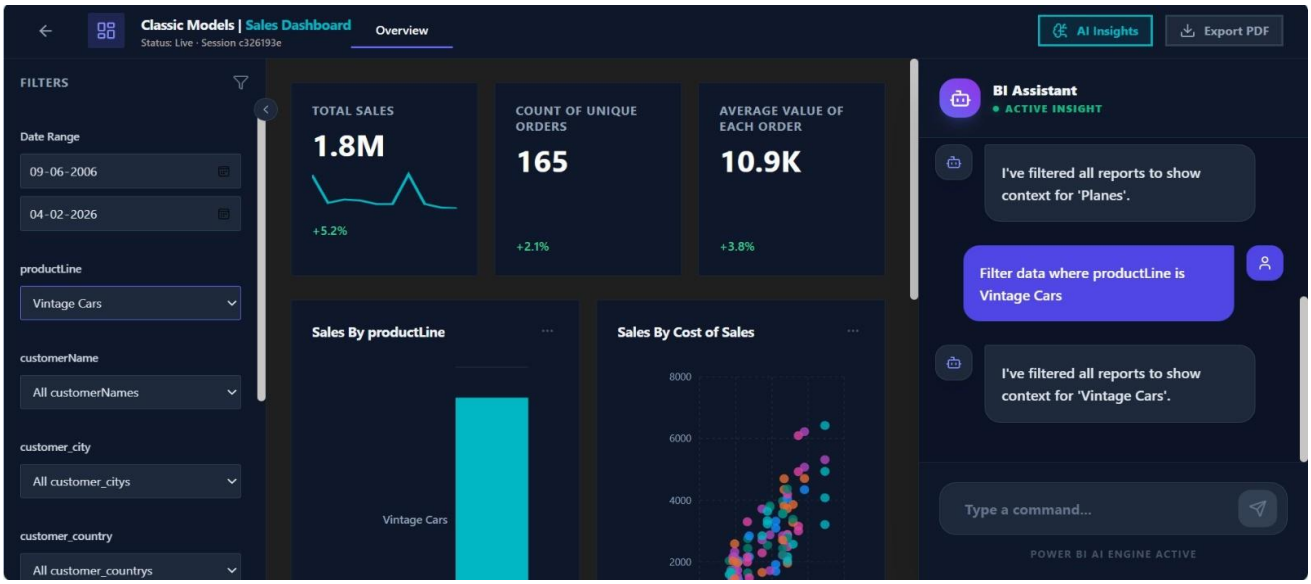


Fig. 5. AutoDashAI dashboard for accuracy comparison.

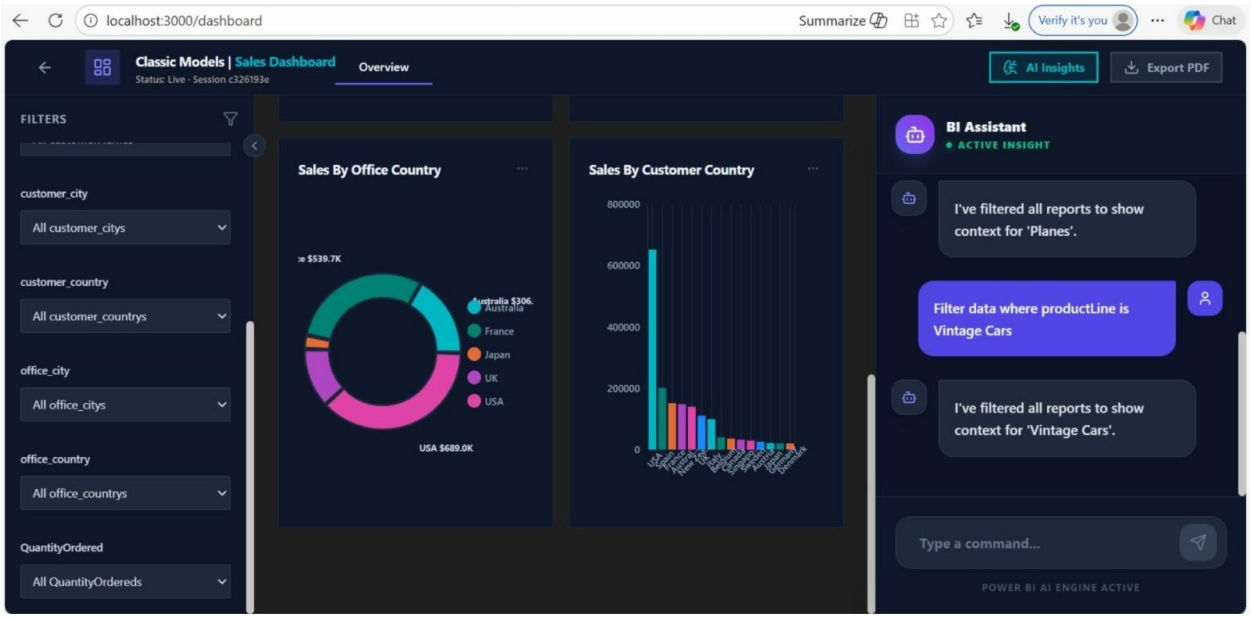


Fig. 6. AutoDashAI dashboard for comparative evaluation.

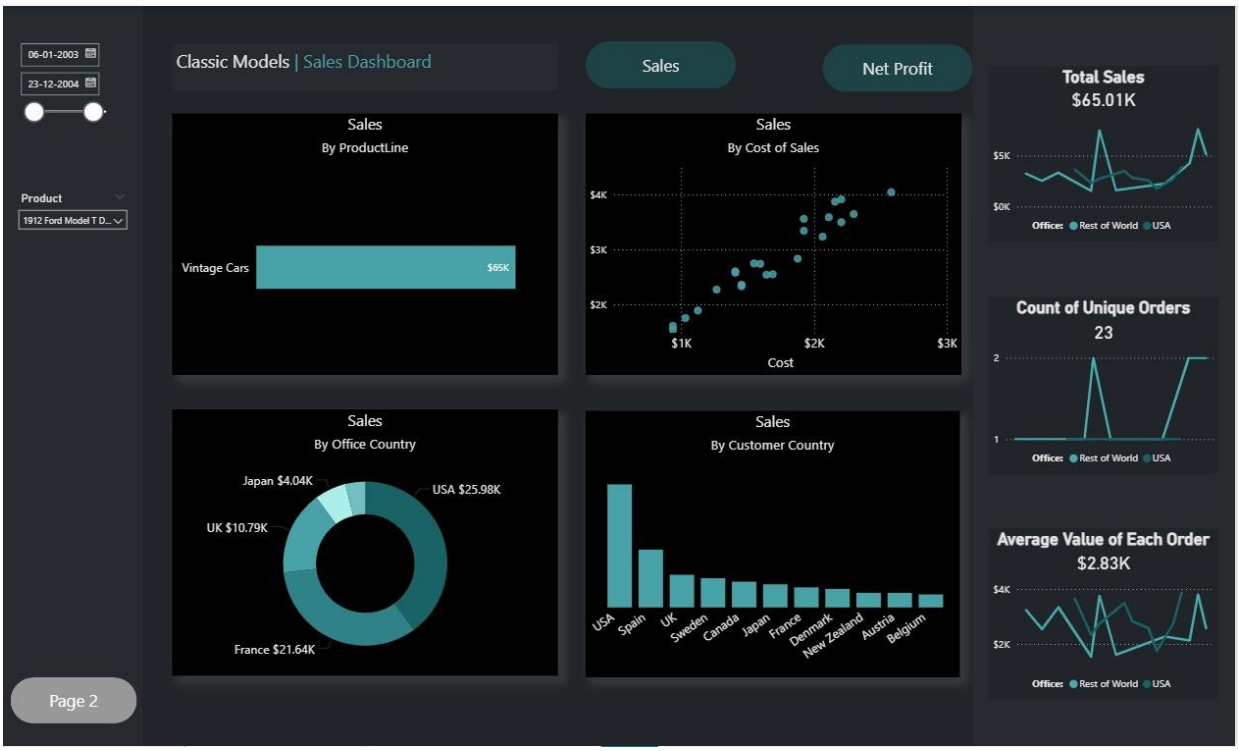


Fig. 7. Same Visualization from Power-BI for Comparison and Accuracy

The insight generating module was able to find statistically significant links by choosing correlations with Pearson coefficients higher than 0.75 for display. To make it easier to understand, high-cardinality category characteristics were automatically grouped together. This cut down on visual clutter and made the dashboard clearer.

D. Hybrid System Reliability

When external LLM services weren't accessible, the deterministic fallback engine was able to carry out around 70% of normal user instructions. This result shows that the hybrid design is strong and may be used in places with poor connection or limited resources.

## 7. Integration with Current Learning Systems

Modern learning environments are relying more and more on data-driven insights to keep track of how well students, staff, and institutions are doing, how engaged they are, and what results they are receiving. Most Learning Management Systems (LMS), including Moodle, Google Classroom, Blackboard, and institutional ERP platforms, on the other hand, create a lot of raw data that people don't utilize much because they don't know how to analyze it and traditional BI tools are too hard to use. AutoDashAI produces interactive dashboards that are particular to schools after it has looked at the data. Faculty can immediately monitor how students are doing over time, locate students who are in trouble, and compare results from other classes or semesters. Administrators may see big-picture information about their institutions, such as attendance patterns, course completion rates, and how resources are being used. Even if users don't know how to use the software, they can ask for specific analyses (such "show attendance vs. grades for first-year students") by using natural language prompts. AutoDashAI supports analytics in multiple regional languages, making it accessible to more students and teachers. The system allows users to design and share dashboards and insights in various languages, enhancing understanding. It also provides comprehensive descriptions of suggested visualizations and patterns, improving confidence and ease of use. This facilitates informed academic decision-making.

Integrating AutoDashAI with existing learning procedures can help schools migrate from static reporting to dynamic, AI- powered insights. This integration makes it easier for teachers to understand data, helps non-technical learners become more data literate, and allows teachers to make swift, evidence-based decisions in modern classrooms.

## 8. Future Directions

The goal of future AutoDashAI development is to advance the system to fully autonomous and adaptive analytics pipelines. This includes real-time ingestion, autonomous schema evolution, self-updated dashboards, and automation of data preparation, visualization, and insight generation. Businesses will be able to monitor dynamic settings without the need for manual configuration thanks to these features.

By merging streaming data sources, such as sensor data and learning activity logs, AutoDashAI can provide real-time analytics and insights. This greatly increases the use of automated dashboards in industries requiring quick response by enabling dynamic graphics and time-sensitive notifications.

AutoDashAI uses visual representations that take ambiguity into account, causal reasoning, and confidence evaluation to simplify complicated information and promote trust. Customers might feel more secure when given clear explanations for suggested charts, patterns, and insights, particularly when making critical decisions. This approach could increase user satisfaction and confidence.

In the future, personalization and AI collaboration are both excellent concepts to think about. For any corporate data collection, adaptive dashboards may change the intricacy of the graphics and the depth of the narratives they tell based on how people use them, their occupations, and their preferences.

Additionally, interactive feedback systems can guide users' analytical study and assist them better their results. This ensures that automated analytics align with what people know about the problem.

In the future, the ethical and responsible aspects of AI will be crucial. Bias detection, fairness-aware analytics, and data processing that protects privacy are necessary for safe deployment. By emphasizing scalability, explainability, customization, and ethical compliance, AutoDashAI may develop into a trustworthy analytics platform for a range of industries and enterprises, making it perfect for practical applications. This approach will boost confidence and trust in AI.

## 9. Conclusion

The study looked at AI-driven automation breakthroughs, specifically dashboard design, improvement, and display. It demonstrates the transition from static to intelligent analytics. The study also addressed issues with natural language interactions, concentrating on upcoming technologies such as agential AI pipelines and massive language models to improve effectiveness. In this regard, AutoDashAI has been offered as a framework that merges data preparation, smart visualization recommen- dation, and AI-based results visualization into a single solution that is prompt-based and no-code. The assistance provided by this tool regarding multi language interactions and results visualization has contributed considerably to its accessibility.

In total, the research findings suggest that the key to developing future analytics platforms will be to achieve the right balance between automation, transparency, scalability and ethical accountability. Even more specifically, AutoDashAI illustrates more inclusive, reliable and user-centered data analytics in real-world applications.

## 10. Acknowledgment

The authors would sincerely like to thank Prof. Vidya Sagvekar of K. J. Somaiya Institute of Technology, Mumbai, for her invaluable guidance and support by providing valuable feedback on the development of the manuscript. The guidance and support of Prof. Vidya Sagvekar have notably contributed to the successful completion of the AutoDashAI project and this Research paper.

The authors are grateful for those from the Department of Artificial Intelligence and Data Science, K.J. Somaiya Institute of Technology, who helped them by giving them the academic environment and provision required to carry out this research. Lastly, the authors would like to express their sincere gratitude to their colleagues and peers for the valuable discussions, insights, and constructive feedback that significantly contributed to the progress and quality of this research.

## References

1. S. K. Gupta, R. Mehta, and A. Patil, "Chat2VIS: Fine-Tuning Data Visualisations using Multilingual Natural Language Text and Pre-Trained Large Language Models," Proc. IEEE Int. Conf. Visual Analytics (VIS), pp. 101–110, 2024.
2. Y. Zhang, L. Sun, and K. Liu, "DAgent: A Relational Database-Driven Data Analysis Report Generation Agent," IEEE Trans. Big Data, vol. 12, no. 3,
3. pp. 456–470, 2024.
4. M. Chen, J. Zhou, and T. Wang, "Dynamic Orchestration of Data Pipelines via Agentic AI," Proc.

IEEE Int. Conf. Data Eng. (ICDE), pp. 311–322, 2024.
5. A. Kumar, V. Sharma, and P. Singh, "Generative AI in Data Science: Applications in Automated Data Cleaning and Preprocessing for Machine Learning Models," IEEE Access, vol. 12, pp. 12,345–12,358, 2024.
6. H. Li, M. Tang, and F. Xu, "Auto Data Visualization via Hierarchical Table Prompting," Proc. ACM/IEEE Joint Conf. Human–Computer Interaction (CHI), pp. 201–212, 2023.
7. R. Davis and E. Miller, "A Review of AI-Powered Data Visualization in Enterprise," IEEE Computer Graphics and Applications, vol. 44, no. 2, pp. 65–77, Mar.–Apr. 2024.
8. S. Banerjee, K. Thomas, and N. Al-Mutairi, "AI-Driven Automation for Big Data Analytics," IEEE Trans. Cloud Computing, vol. 13, no. 4, pp. 842–856, Jul.–Aug. 2024.
9. J. Park and H. Yoon, "HAIChart: Human and AI Paired Visualization System," Proc. IEEE Conf. Visual Analytics Science and Technology (VAST), pp. 121–130, 2024.
10. X. Wu, Z. Zhou, and M. Lin, "AdaVis: Adaptive and Explainable Visualization Recommendation for Tabular Data," IEEE Trans. Visualization and Computer Graphics (TVCG), vol. 30, no. 1, pp. 215–229, Jan. 2023.
11. L. Chen, Y. Zhao, and D. Huang, "AI-Enhanced Data Visualization: Transforming Complex Data into Actionable Insights," IEEE Access, vol. 12, pp. 45,210–45,225, 2024.