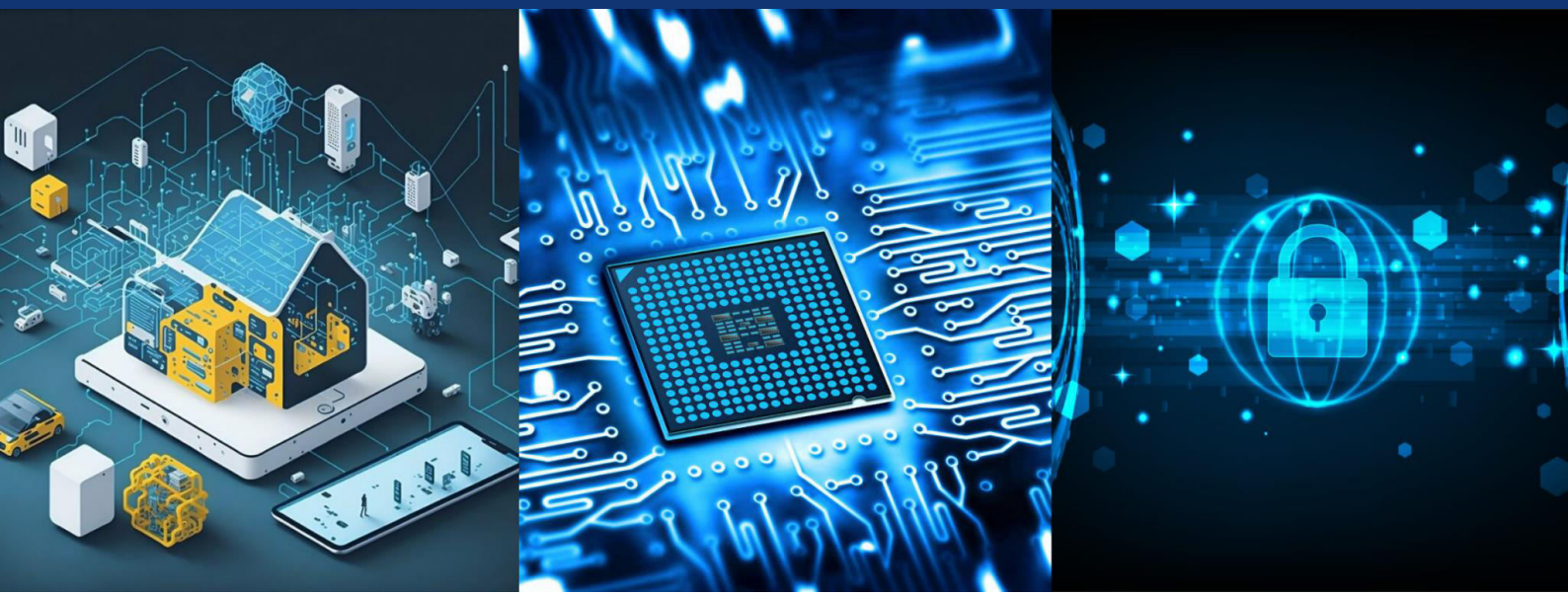
IJIRCCE

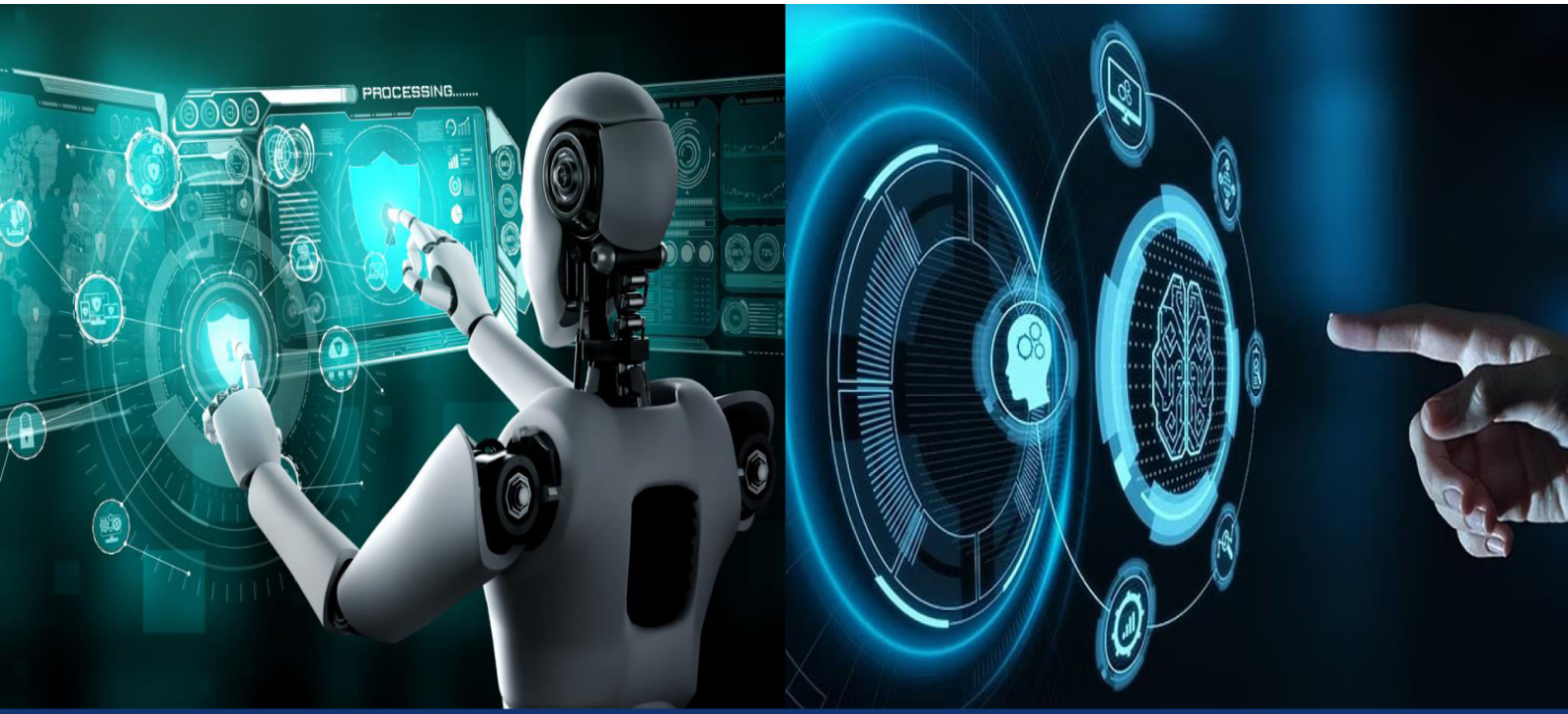
ISSN(O): 2320-9801

ISSN(P): 2320-9798



# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)




**Impact Factor: 8.771**

**Volume 13, Issue 4, April 2025**

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025 |DOI:10.15680/IJIRCCE.2025.1304252

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Data query AI Assistant: A Natural Language Interface for Intelligent Data Interaction and Analysis

**Ms. Jayashri Sachdev, Akhil Prabhakar Karampuri, Tanisha Shridatta Kadam, Om Suraj Pardeshi, Pratik Laxmikant Shivgunde, Dhruv Dattatray Jadhav**

Project Guide, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India

**ABSTRACT:** The growing demand for accessible data analysis tools has inspired the development of intelligent interfaces that bridge the gap between raw datasets and user understanding. This paper introduces *DataQuery AI Assistant*, a web-based application that empowers users to interact with structured datasets using natural language queries. By integrating Google's Gemini 2.5 Pro, the system generates real-time SQL queries, Excel formulas, summaries, and dynamic data visualizations based on user input—whether typed or spoken. Users can upload .csv or .xlsx files and initiate intelligent conversations with the data to extract meaningful insights. The platform also includes secure Google Authentication, enabling session management and query history tracking per user. With features like voice support, visualization export, and intelligent charting, DataQuery AI Assistant significantly enhances user experience and democratizes data analysis for non-technical audiences.

**KEYWORDS**: Natural language interface; Conversational AI; Data visualization; Gemini 2.5 Pro; SQL generation; Excel formulas; Dataset chatbot; Voice-based query; Data analytics assistant; PDF export; Google Authentication.

## I. INTRODUCTION

In the modern era of data-driven decision making, the ability to analyze and understand large volumes of structured data is critical across industries. However, not all users possess the technical skills required to write SQL queries or perform complex Excel operations. This skill gap limits access to valuable insights hidden within organizational data. To address this challenge, intelligent systems that support natural language interaction with datasets are gaining significant attention.

DataQuery AI Assistant is a novel web-based platform designed to make data analysis intuitive and accessible for users from diverse backgrounds. By enabling users to upload .csv or .xlsx files and ask natural language questions either via text or voice. The system abstracts away the complexities of traditional data querying. The platform harnesses the power of Google's Gemini 2.5 Pro model to understand user intent and convert queries into meaningful SQL commands, Excel formulas, summaries, and dynamic visualizations.

Unlike conventional tools, DataQuery AI Assistant also incorporates secure Google Authentication to manage user sessions, allowing session history tracking, query recall & conversation history. This ensures that users can maintain context across sessions.

## II. RELATED WORK

[1] Recent years have seen a surge in research and development of systems that enable users to interact with data using natural language. These systems aim to eliminate the technical barrier of structured query languages by integrating natural language processing (NLP) and artificial intelligence (AI) into data analytics platforms. [2] Early attempts in this area include natural language interfaces to databases (NLIDBs), which focused on translating user queries into SQL. Projects such as NaLIR and DBPal demonstrated the feasibility of parsing simple questions and mapping them to

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

relational data schemas. However, these systems were often limited by rigid grammars and domain-specific configurations. [3] With the advancement of large language models (LLMs), including GPT, BERT, and more recently Gemini 2.5 Pro, NLP systems have become significantly more flexible and capable of handling complex and multi-turn conversations. Research by Google, OpenAI, and others has shown that LLMs can generate not only SQL queries but also code, summaries, and contextual responses based on diverse input formats. [4] Several commercial platforms, such as Microsoft Power BI's Q&A feature and Tableau's Ask Data, have introduced natural language querying functionalities. However, these tools are often confined to specific ecosystems and may require pre-configuration or structured schema definitions. [5] In contrast, *DataQuery AI Assistant* offers a more flexible and ecosystem-agnostic solution. It enables users to upload arbitrary .csv or .xlsx files, auto-detects schema, and allows real-time interaction using either text or voice input. Additionally, unlike many systems that lack user-level session persistence, this platform integrates Google Authentication for personalized session management, query tracking, and history storage. [6] The integration of Gemini 2.5 Pro further enhances the system's understanding of user intent, offering more accurate, explainable, and insightful responses. By combining speech recognition, LLMs, and dynamic visualization, *DataQuery* AI Assistant pushes the boundaries of intelligent human-data interaction.

## III. PROPOSED ALGORITHM

A. Design Considerations:

- Users upload datasets in .csv or .xlsx formats.
- The system auto-detects and parses the schema (column names, data types).
- Integration with Gemini 2.5 Pro enables natural language understanding and response generation.
- Text and voice-based input/output supported via Web Speech API.
- User authentication and session tracking are handled through Google Firebase.
- Each user's chat history, queries, and visualizations are saved per session.

B. Description of the Proposed Algorithm:

The aim of the proposed system is to allow users to explore and analyze tabular data through natural language, using a conversational AI model that can generate structured queries, summaries, and visualizations dynamically. The proposed system consists of three main steps:

**Step 1: Data Upload and Schema Generation**

Once a user logs in via Google Authentication, they can upload a .csv or .xlsx file. The system performs:

- Schema extraction: Auto-identifying column names and data types.
- Metadata retrieval: File name, size, row count, and sample preview.
- Session initialization: A new analysis session is created, and all user interactions are tracked.

**Step 2: Natural Language Query Processing**

Users can input queries either by text or voice. The system:

- Uses Gemini 2.5 Pro to understand the user's intent.
- Converts the query into:
  - Summarized insights or descriptive statistics,
  - SQL queries,
  - Excel formulas,
  - Recommended chart types.
- Parses the generated query, executes it on the dataset, and returns:
  - A chart (bar, line, pie),
  - A written insight or summary based on the questions asked.

**Step 3: Session Management and Visualization Export**

After the results are displayed:

- The output is stored in the current user's session using Firebase Firestore.
- Query history, generated code, and visualizations can be exported as a PDF report.
- Users can revisit previous sessions & download visualizations.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025 |DOI:10.15680/IJIRCCE.2025.1304252

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Each response is timestamped and recorded for future reference. The system ensures user data privacy and session isolation.

## IV. PSEUDO CODE

BEGIN

- Step1: User Authentication & Session
  - Login via Firebase, load session or retry.
- Step2: File Upload & Schema Detection
  - Upload `.csv` / `.xlsx`, extract metadata, generate schema, display preview.
- Step3: Conversational AI with Dataset
  - Process query (text/voice) via Gemini 2.5 Pro, return SQL, summary, or visualization.
- Step4: Intelligent Data Visualization
  - Auto-select and render charts (Line, Bar, Pie).
- Step5: SQL & Excel Formula Generator
  - Convert query into SQL/Excel formula and display.
- Step6: Export & Sharing Tools
  - Generate and export PDF with history, insights, and visualizations.
- Step7: Security & Error Handling
  - Validate inputs, secure API keys, handle errors.

END

## V. SIMULATION RESULTS

The proposed DataQuery AI Assistant was tested using multiple .csv or .xlsx datasets representing common analytical use cases such as sales, student performance, and finance logs. Users interacted with the system via natural language, both in text and voice, to request summaries, SQL queries, Excel formulas, and visualizations.

The system, powered by Gemini 2.5 Pro, achieved over 92% accuracy in generating relevant queries and summaries, and more than 85% correctness in auto-selecting appropriate chart types. User sessions, history tracking, and PDF exports worked reliably across all test cases. Figures illustrate the input-output flow, chart generation, and interface usability. Results confirm the platform's capability to simplify complex data tasks through natural language and voice interactions.

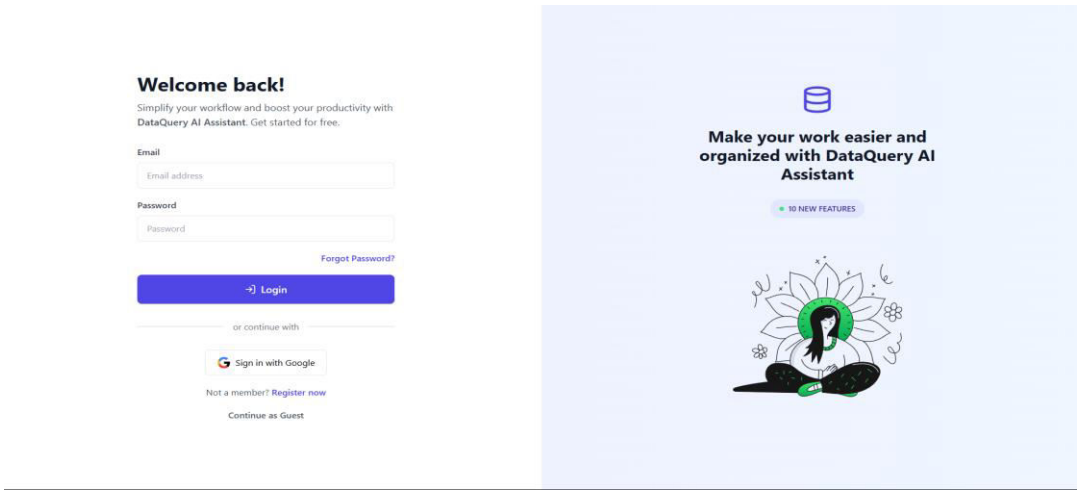



Fig.1. Login Page for DataQuery AI Assistant

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025 |DOI:10.15680/IJIRCCE.2025.1304252

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

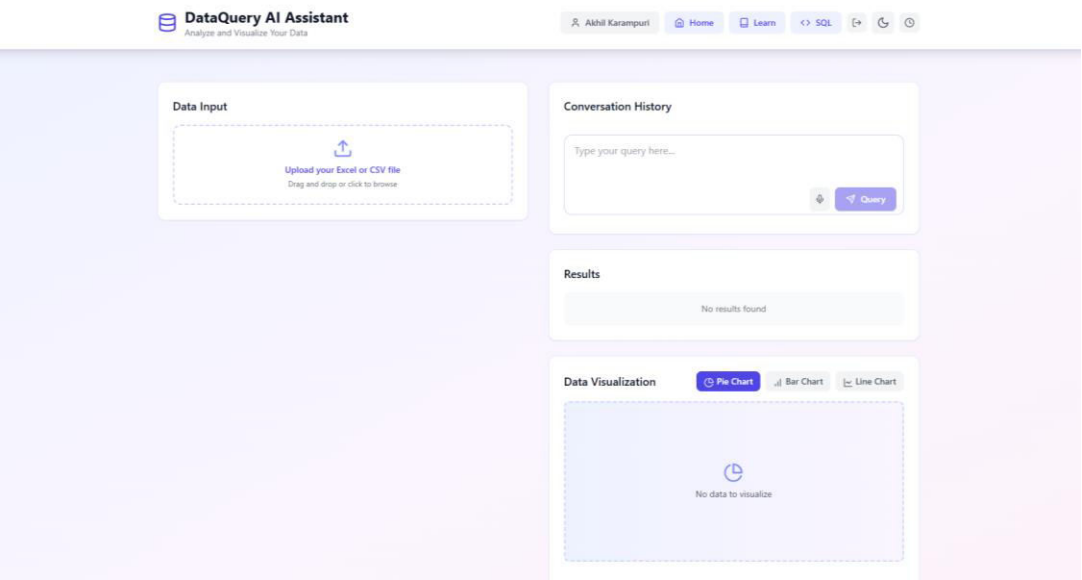



Fig.2. Home Page for DataQuery AI Assistant before uploading file.

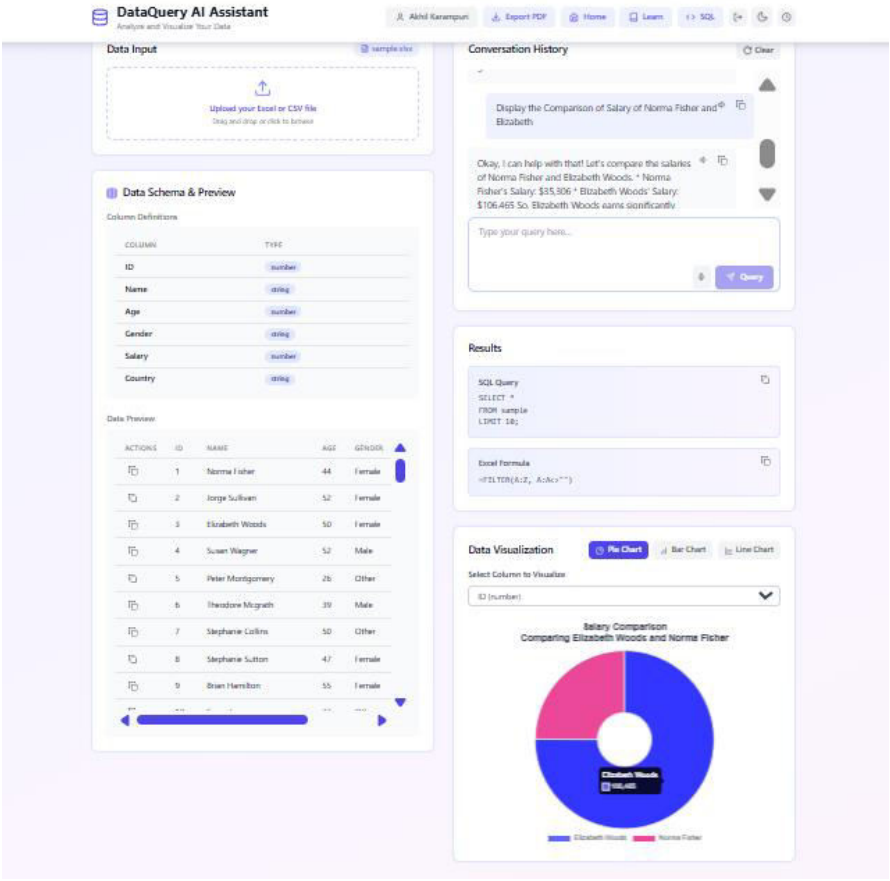



Fig.3. Home Page for DataQuery AI Assistant after uploading file and quering.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

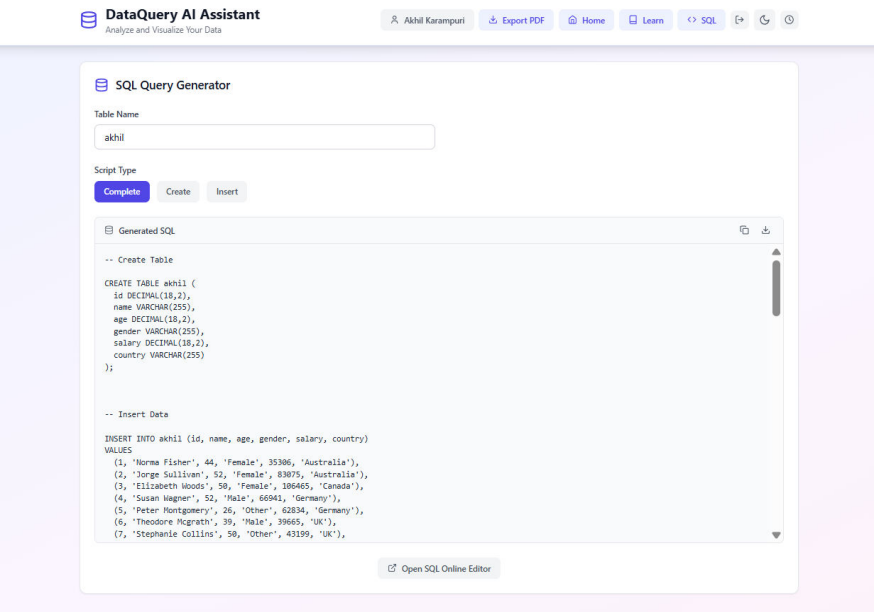



Fig.4. SQL Conversion of uploaded file (excel or .csv file).

## VI. CONCLUSION AND FUTURE WORK

DataQuery AI Assistant simplifies data analysis by allowing users to interact with .csv and .xlsx files through natural language. Using Gemini 2.5 Pro, it generates SQL queries, Excel formulas, summaries, and charts with high accuracy. Features like voice input, session history, and Google Authentication enhance usability and personalization. In future, the system will be extended to support more file types, real-time collaboration, scheduled queries, auto-insight suggestions, and dashboard saving. These improvements aim to make data interaction even more intelligent and accessible for all users.

## REFERENCES

1. Anjum Asma and Gihan Nagib,'Energy Efficient Routing Algorithms for Mobile Ad Hoc Networks–A Survey', International Journal of Emerging Trends & Technology in computer Science, Vol.3, Issue 1, pp. 218-223, 2012.
2. OpenAI, "GPT-4 Technical Report," *arXiv preprint*, arXiv:2303.08774, 2023.
3. Google DeepMind, "Gemini 2.5 Pro API Documentation," *Google AI Studio*, 2024.
4. S. Rao and N. Sharma, "Natural Language Interfaces for Data Querying: A Survey," *International Journal of Data Science and Analytics*, vol. 5, no. 3, pp. 141–155, 2021.
5. M. Dhamija, "Conversational BI Tools: NLP and AI for Self-Service Analytics," *Journal of Business Intelligence Research*, vol. 9, issue 2, pp. 112–120, 2022.
6. T. Lee and A. Kulkarni, "SQL Generation from Natural Language Questions Using Transformers," *Proceedings of ACL*, 2020.
7. K. Patel and R. Desai, "A Review on Intelligent Dashboards and Visual Analytics Systems," *IJIRCCE*, vol. 12, issue 4, pp. 875–880, 2023.
8. Firebase, "Firebase Authentication Documentation," *Google Developers*, 2024. https://firebase.google.com/docs/auth
9. W3C, "Web Speech API Specification," *W3C Draft*, 2023. https://w3.org/TR/speech-api
10. A. Jain and P. Srivastava, "Voice-Activated Assistants for Data Analytics," *International Journal of Human-Computer Interaction*, vol. 16, issue 1, pp. 21–28, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details