International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

EXCEL FORMULAS AND CLEANUP COPILOT

[1] A Satyanarayana, [2] Manan Rout, [3] K Malleshwari, [4] B Kalyan, [5] M Vamshi
[1]Assistant Professor, [2345]Students
Department of CSE(AIML)
Siddhartha Institute of Technology & Sciences, Narapally
drsatyanarayanaakella_cse@siddhartha.co.in, 23TQ1A66C1@siddhartha.co.in,
23TQ1A6670@siddhartha.co.in, 23TQ1A66C5@siddhartha.co.in, 23TQ1A66A4@siddhartha.co.in

Abstract

The **Excel Formulas and Cleanup Copilot** is an Artificial Intelligence-based productivity system designed to help users create spreadsheet formulas, clean datasets, identify errors, and improve the overall quality of Excel worksheets. Spreadsheet applications are widely used in education, business, finance, administration, and data analysis. However, users often face difficulties when working with complex formulas, inconsistent data, duplicate records, missing values, and formatting problems.

The proposed system allows users to upload an Excel file or enter a description of the task they want to perform. The AI analyzes the spreadsheet structure, identifies columns, data types, patterns, and existing formulas, and understands the user's natural-language request. Based on this analysis, the system can generate appropriate Excel formulas such as SUM, AVERAGE, IF, COUNTIF, XLOOKUP, INDEX-MATCH, and other commonly used functions.

A major feature of the system is automated spreadsheet cleanup. The Copilot can detect duplicate records, empty cells, inconsistent capitalization, unwanted spaces, incorrect data formats, invalid values, and inconsistent date or number representations. It can recommend appropriate cleaning operations and allow users to review the proposed changes before applying them. This reduces repetitive manual work and improves data consistency.

The system also provides formula explanations and error assistance. When a formula produces an error, the AI can identify possible causes and suggest a corrected formula. Users can describe their requirements in simple language, such as "calculate the total sales for each employee" or "find customers whose sales are greater than 50,000," and the system can translate these requirements into suitable spreadsheet formulas.

Overall, the **Excel Formulas and Cleanup Copilot** aims to make spreadsheet work faster, easier, and more reliable. It can be useful for students, employees, analysts, accountants, administrators, and business users who frequently work with Excel data. Future enhancements can include direct Excel integration, automated chart generation, advanced data visualization, natural-language spreadsheet editing, formula optimization, and intelligent report generation.

I. INTRODUCTION

Microsoft Excel and other spreadsheet applications are widely used for storing, organizing, calculating, and analyzing data. Users rely on spreadsheets for tasks such as financial calculations, attendance management, sales analysis, inventory management, budgeting, and reporting. Although spreadsheets provide powerful functionality, users need sufficient knowledge of formulas, formatting, and data-cleaning techniques to use them effectively.

Writing complex formulas can be challenging, particularly for beginners and users who do not frequently work with advanced Excel functions. A small mistake in a cell reference, function name, range, or condition can produce an incorrect result. Understanding and correcting spreadsheet errors can also require significant time. Users often search online or manually experiment with formulas to find a solution.

Spreadsheet data can also contain many quality problems. Duplicate records, missing values, extra spaces, inconsistent capitalization, incorrect date formats, and mixed data types can make analysis unreliable. Traditional cleanup requires users to manually inspect rows and columns or use separate Excel features. When datasets are large, identifying every inconsistency becomes difficult and time-consuming.

Artificial Intelligence and Natural Language Processing provide an opportunity to simplify spreadsheet operations. Instead of remembering complicated formula syntax, users can describe their requirements in natural language. An AI system can understand the request, inspect the spreadsheet structure, generate a suitable formula, explain it, and identify potential errors. AI can also analyze data patterns and recommend appropriate cleanup operations.

The **Excel Formulas and Cleanup Copilot** combines formula generation, formula explanation, error detection, data cleaning, and spreadsheet analysis in one intelligent platform. The goal is not to replace the user's control over the spreadsheet but to provide intelligent assistance. Users can review suggested formulas and cleanup actions before applying them, making the system useful for both beginners and experienced spreadsheet users.

II. LITERATURE SURVEY

Traditional spreadsheet applications provide a wide range of functions for calculations, data organization, sorting, filtering, and formatting. Excel includes hundreds of formulas and built-in tools that allow users to perform complex operations. However, users generally need to know the correct syntax and understand spreadsheet functions before they can use these features efficiently.

Spreadsheet error-detection techniques have been studied to identify formula inconsistencies, incorrect references, circular dependencies, and unusual calculation patterns. Built-in spreadsheet auditing tools can help users trace formulas and identify certain errors. However, these tools generally require users to understand spreadsheet logic and may not provide natural-language explanations for beginners.

Data-cleaning techniques are commonly used in database systems and data-analysis applications. Operations such as removing duplicates, handling missing values, standardizing formats, and correcting inconsistent data are fundamental preprocessing

steps. Traditional approaches require predefined rules or manual configuration, which can become difficult when datasets contain different types of inconsistencies.

Natural Language Processing has enabled systems to translate human-language instructions into structured operations. In spreadsheet environments, natural-language interfaces can allow users to describe calculations or data operations without manually writing formulas. Such systems can improve accessibility for users who understand the business requirement but have limited knowledge of spreadsheet syntax.

Recent advances in **Generative AI and Large Language Models** have created new possibilities for spreadsheet assistance. AI systems can generate formulas, explain calculations, identify potential errors, summarize data, and recommend transformations based on natural-language requests. However, generated formulas may sometimes be incorrect, so validation against the spreadsheet structure and user review are important.

Existing research and intelligent spreadsheet tools demonstrate the benefits of combining AI with spreadsheet analysis. However, there is still a need for systems that integrate formula generation, error explanation, data-quality analysis, cleanup recommendations, and user-controlled execution. The proposed **Excel Formulas and Cleanup Copilot** addresses this need by combining these capabilities into a single AI-powered workflow.

III. SYSTEM ANALYSIS

The **Excel Formulas and Cleanup Copilot** is designed to assist users in creating formulas and improving spreadsheet data quality. The user first uploads an Excel workbook or selects a worksheet through the system interface. The spreadsheet-analysis module reads the workbook structure, including sheets, columns, cell values, existing formulas, and data types. The system identifies patterns and detects potential data-quality problems such as duplicates, missing values, extra spaces, and inconsistent formats. The user can then enter a natural-language request describing the required calculation or cleanup operation. An NLP module interprets the request and identifies the required columns, conditions, functions, and operations. The AI formula engine generates a suitable Excel formula based on the spreadsheet structure. A validation module checks references, ranges, syntax, and compatibility with the selected worksheet. The cleanup engine recommends appropriate transformations for detected data-quality issues. The system explains formulas and cleanup actions in simple language. Users can preview proposed changes and approve, reject, or modify them. Finally, the system applies approved operations and provides the cleaned workbook and generated formulas as output.

Existing System

Traditional Excel workflows require users to manually create formulas and perform data-cleaning operations. Users need to understand Excel functions and correctly enter cell references, ranges, conditions, and lookup operations. When formulas become complex, users may depend on online tutorials, documentation, or trial-and-error methods. Spreadsheet errors can be difficult to identify when a workbook contains many sheets and formulas. Data cleanup is also frequently performed using manual filtering, sorting, Find and Replace, formatting tools, and built-in Excel features. Users

may need to inspect large datasets to identify duplicates, missing values, and inconsistent formats. Different cleanup tasks may require different Excel tools or additional software. Existing spreadsheet features generally do not understand detailed natural-language requirements in a conversational way. They also provide limited explanations of why a formula is incorrect or how a particular cleanup operation should be performed. As a result, beginners may find advanced spreadsheet tasks difficult and experienced users may spend considerable time on repetitive operations.

Disadvantages of Existing System

- Requires knowledge of Excel formulas and syntax.
- Complex formulas can be difficult to create manually.
- Formula errors may be difficult to identify.
- Users often depend on external tutorials or documentation.
- Data cleanup can require significant manual effort.
- Difficult to detect all inconsistencies in large datasets.
- Limited natural-language interaction.
- Limited explanation of formula errors.
- Duplicate and missing data may require manual inspection.
- Formatting inconsistencies can be overlooked.
- Repetitive cleanup tasks consume time.
- Different spreadsheet operations may require separate tools.

Proposed System

The proposed **Excel Formulas and Cleanup Copilot** uses Artificial Intelligence and Natural Language Processing to simplify spreadsheet calculations and data-cleaning tasks. The user can upload an Excel workbook and describe the required operation using natural language. The system analyzes the workbook structure and identifies sheets, columns, data types, formulas, and data patterns. An NLP engine interprets the user's request and maps it to the relevant spreadsheet columns and operations. The AI formula generator creates suitable Excel formulas based on the workbook context. A validation engine checks formula syntax, cell references, ranges, and logical conditions before presenting the result. The system also analyzes the dataset for duplicates, missing values, inconsistent text, extra spaces, invalid formats, and unusual values. A cleanup recommendation engine suggests suitable transformations for the detected problems. Users can preview every suggested formula and cleanup action before applying changes. The Copilot provides simple explanations so users understand what each formula or transformation does. Users can approve, reject, or modify individual suggestions. Finally, the system produces an updated workbook containing the approved formulas and cleaned data.

Advantages of Proposed System

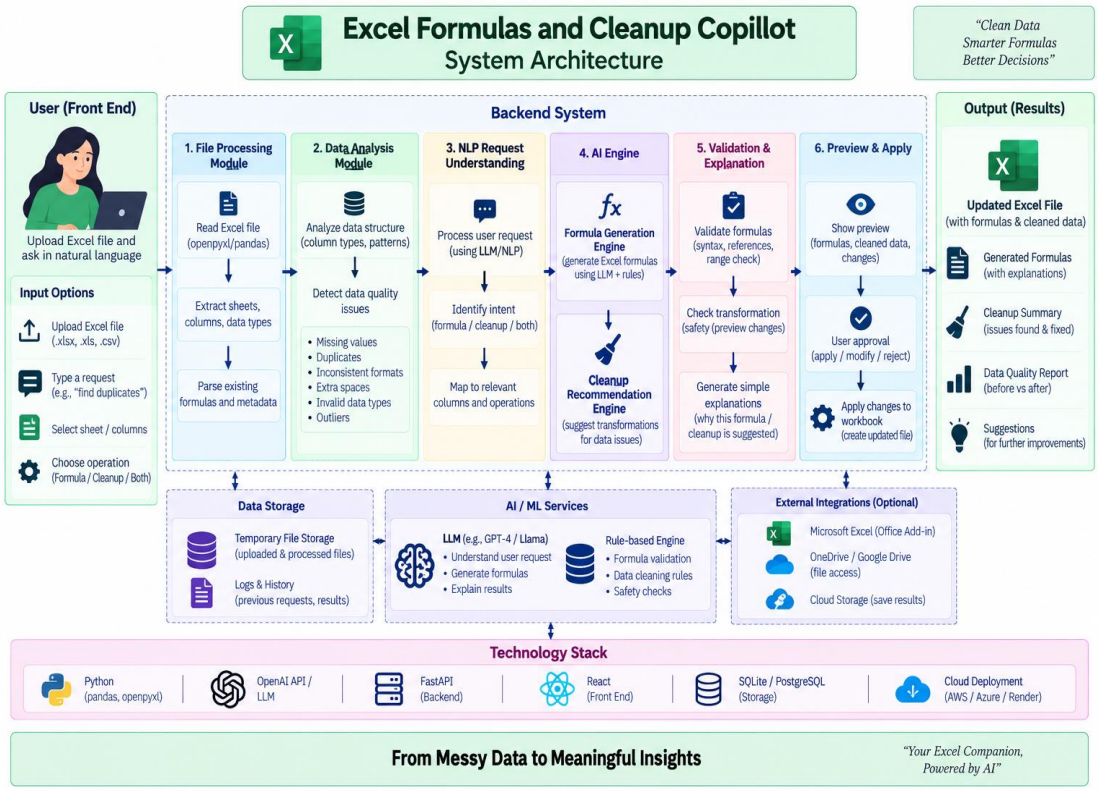
- Generates Excel formulas from natural-language instructions.
- Reduces the need to remember complex formula syntax.
- Explains formulas in simple language.
- Detects common spreadsheet errors.
- Identifies duplicate records automatically.
- Detects missing and inconsistent values.

- Helps standardize dates, numbers, and text formats.
- Provides intelligent cleanup recommendations.
- Allows users to preview changes before applying them.
- Saves time on repetitive spreadsheet operations.
- Supports beginners as well as experienced users.
- Combines formula assistance and data cleanup in one system.

IV. METHODOLOGY

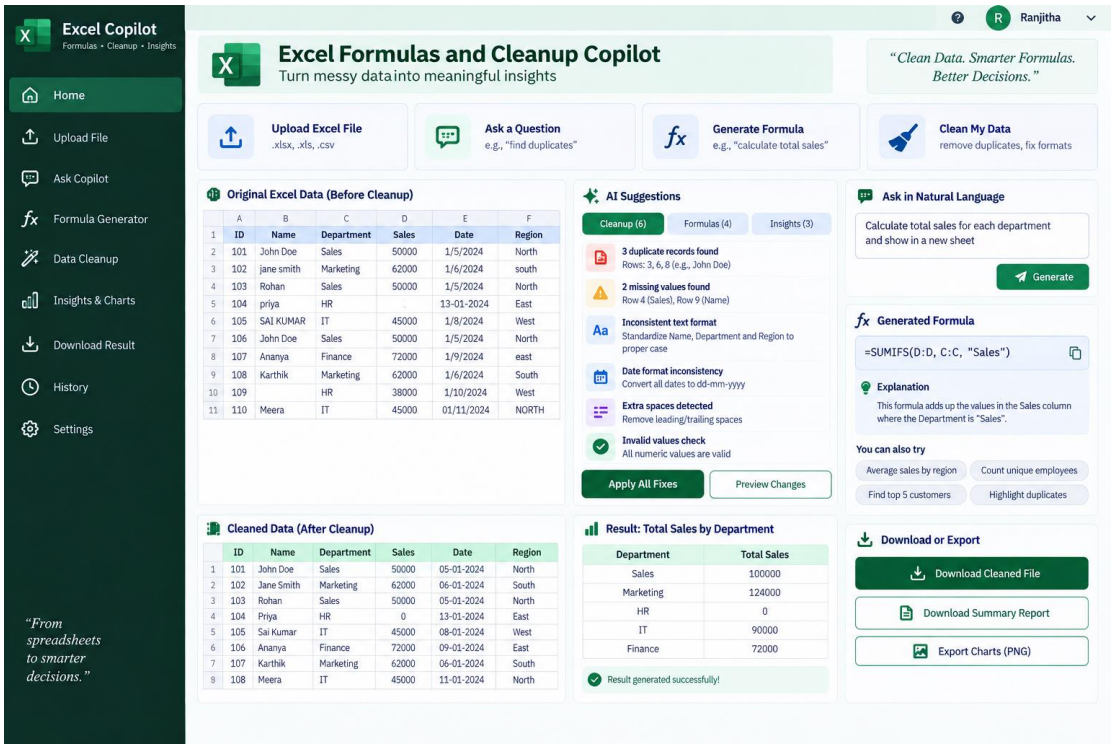
The methodology begins when the user uploads an Excel workbook through the Copilot interface. The system validates the workbook and extracts worksheet names, columns, cell values, formulas, data types, and formatting information. A spreadsheet-analysis module examines the dataset to identify patterns and potential data-quality problems. The user then provides a natural-language instruction describing the desired calculation, formula, or cleanup operation. The NLP module analyzes the instruction and identifies relevant columns, conditions, functions, and transformations. The AI formula-generation engine creates a suitable Excel formula based on the workbook structure. A validation module checks formula syntax, cell references, ranges, and logical consistency. The cleanup engine detects duplicates, missing values, unwanted spaces, inconsistent capitalization, and formatting differences. The system generates recommended cleanup operations and explains the reason for each recommendation. Users can preview the generated formulas and proposed changes before approving them. Approved operations are applied to the workbook while maintaining the original data where possible. Finally, the system generates the updated spreadsheet along with formula explanations, cleanup summaries, and detected-error reports.

SYSTEM ARCHITECTURE



The Excel Formulas and Cleanup Copilot consists of multiple interconnected modules that work together to provide intelligent spreadsheet assistance. The User Interface allows users to upload Excel workbooks and enter natural-language requests. The Workbook Processing Module reads worksheets, cell values, formulas, formatting, and metadata. The Spreadsheet Structure Analyzer identifies columns, data types, relationships, and existing formula patterns. The Data Quality Analyzer detects duplicates, missing values, inconsistent text, invalid formats, and unusual values. The NLP Request Understanding Module converts the user's natural-language instruction into structured spreadsheet operations. The AI Formula Generation Engine creates formulas according to the workbook context and user requirements. The Formula Validation Engine checks syntax, references, ranges, and logical conditions. The Cleanup Recommendation Engine suggests appropriate data transformations. The AI Explanation Module explains formulas, errors, and cleanup recommendations in simple language. The Preview and Approval Module allows users to review proposed changes before execution. The Workbook Transformation Engine applies approved formulas and cleanup operations. Finally, the Output and Reporting Module provides the updated Excel workbook, formula explanations, cleanup summary, and error report.

V. RESULT AND OUTPUT



VI. CONCLUSION

The Excel Formulas and Cleanup Copilot provides an intelligent and efficient solution for simplifying spreadsheet operations. The system helps users generate formulas, identify spreadsheet errors, clean inconsistent data, and perform common Excel tasks without requiring advanced knowledge of spreadsheet functions.

By combining Artificial Intelligence, Natural Language Processing, data analysis, and formula-generation techniques, the system allows users to describe their requirements

in simple natural language. It can generate suitable formulas, explain their purpose, identify potential errors, and recommend cleanup operations such as removing duplicates, handling missing values, correcting formatting, and standardizing data.

The system also provides a preview and approval mechanism, allowing users to review suggested formulas and data transformations before applying them. This improves reliability and gives users control over important spreadsheet changes. The generated results and cleanup reports make it easier to understand how the original dataset was modified.

Overall, the Excel Formulas and Cleanup Copilot makes spreadsheet management faster, simpler, accurate, and user-friendly. It can be useful for students, business professionals, accountants, analysts, and other users who regularly work with Excel data. Future enhancements can include direct Microsoft Excel integration, advanced chart generation, automated report creation, voice-based spreadsheet commands, formula optimization, intelligent data insights, and integration with cloud storage platforms.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.