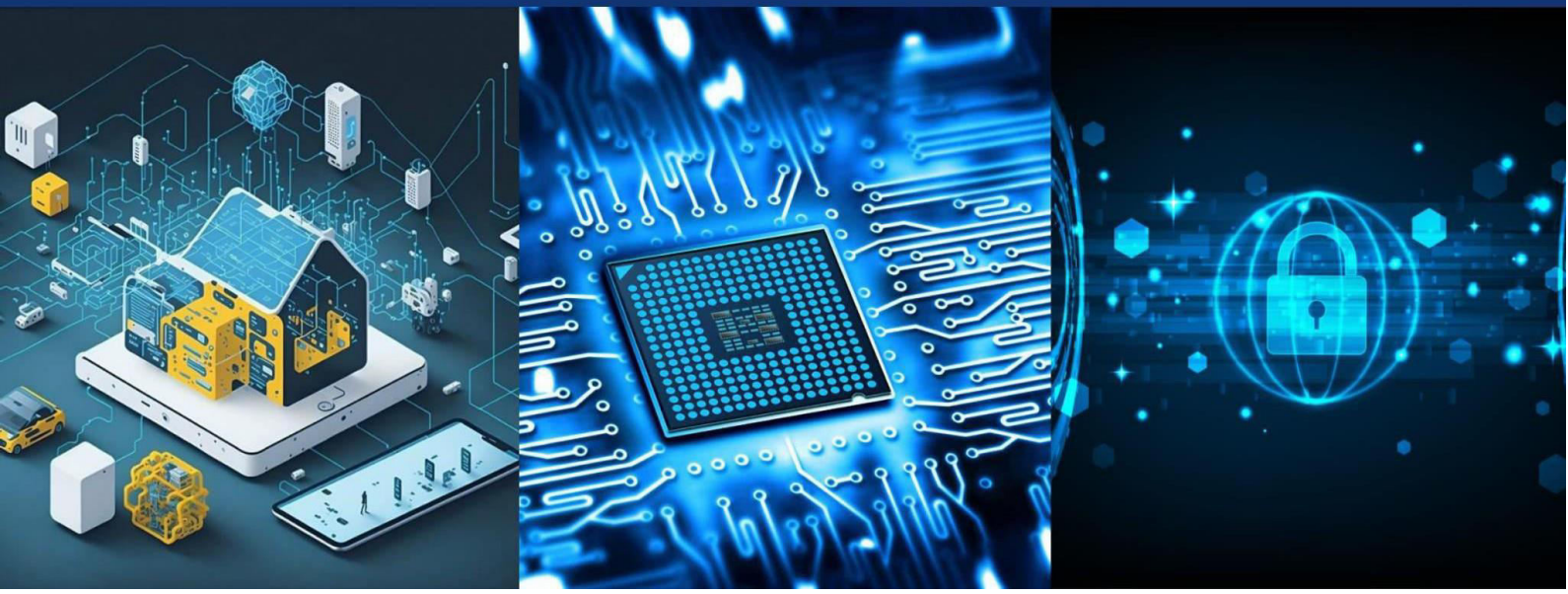
IJIRCCE

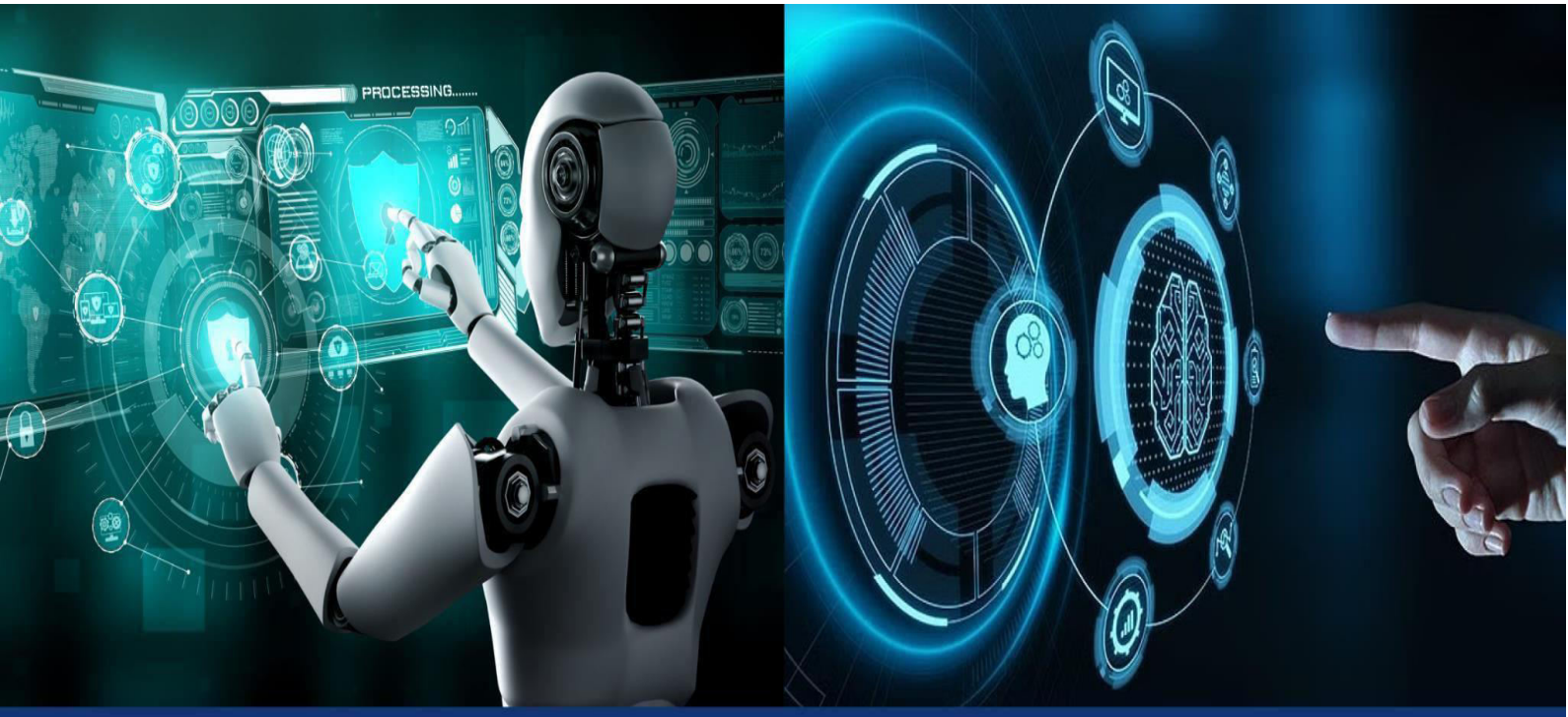
ISSN(O): 2320-9801

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 12, December 2025

© 2025 IJIRCCE | Volume 13, Issue 12, December 2025| DOI:10.15680/IJIRCCE.2025.1312045

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)

AI-Based Automated File Organizer

Prof. Anu C S[1], Amogh P Puthanikar[2], Vibha G S[2], Vineeth S S[2]

Assistant Professor, Department of CS&E, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[1]

U.G. Student, Department of CS&E, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[2]

ABSTRACT: The rapid growth of digital data has made manual file organization increasingly inefficient, time-consuming, and prone to human error. Users often accumulate thousands of unorganized files across devices, leading to clutter and reduced productivity. This project proposes an AI Based Automated File Organizer designed to intelligently analyze, categorize, and structure files using both rule-based techniques and artificial intelligence. The system employs file-type detection, content extraction, and semantic understanding to generate meaningful categories for documents, images, and media files. PDF content is processed using a local large language model to identify themes and generate summaries. The organizer operates through a user friendly Electron.js desktop application and ensures privacy by performing all AI processing locally. Mixed analysis combines type-based and AI-based methods, offering high accuracy and flexibility. The result is an efficient, scalable, and intelligent file-management solution that significantly reduces digital clutter and improves file retrieval. This project demonstrates the practical application of AI in enhancing digital organization and personal productivity.

KEYWORDS: Automated File Management, Semantic File Categorization, Content Extraction, PDF Analysis, Local Large Language Model, Electron.js Application, Mixed Sorting Approach, Digital Clutter Reduction, Intelligent File Retrieval, AI Content Analysis.

I. INTRODUCTION

The rapid growth of digital data has led to cluttered file systems, where users struggle to manage diverse documents such as PDFs, Word files, images, and spreadsheets. Traditional manual organization is time-consuming, inconsistent, and inefficient, causing frequent file-loss, redundancy, and reduced productivity. Existing file managers rely on basic metadata and cannot understand the semantic content of documents. To address these challenges, we propose an AI-based, locally-run automated file organizer that categorizes files using content-aware analysis rather than simple file names or extensions. The system extracts text using tools like PyMuPDF, python-docx, and Tesseract, and applies NLP with spaCy along with clustering algorithms for semantic grouping. Large language models such as LLaMA 3.2, running locally via Ollama, enable intelligent classification and summarization while preserving user privacy. The application supports multimodal inputs, real-time folder monitoring, and dynamic directory creation. Built with Electron.js, it offers a modern, cross-platform interface for intuitive interaction. This project introduces a secure, efficient, and intelligent solution that transforms digital file management into an automated, context-driven process.

II. LITERATURE SURVEY

Li et al. proposed a multimodal document-organization system that integrates text and image embeddings to improve heterogeneous file classification accuracy. The approach significantly enhances mixed-format clustering but suffers from high computational overhead [1].

Zhang et al. utilized transformer-based sentence embeddings to achieve semantically meaningful document clusters. While effective for topic grouping, their method is computationally expensive for large datasets [2].

The LLaMA research demonstrates that smaller large language models can run locally, enabling private, on-device document summarization and classification. This local-first capability makes LLaMA suitable for privacy-sensitive environments without cloud dependency [3].

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)

Yang et al. introduced a CNN-based image-similarity pipeline for grouping scanned pages and visually rich documents. Their approach is highly effective when text extraction is limited or unreliable, such as in scanned files [4].

Raza et al. investigated embedding-based clustering using K-means for organizing text-heavy document collections. Although the method performs well on clean text, it struggles with OCR noise and multimodal documents [5].

Farzana et al. examined dimensionality-reduction methods such as PCA and SVD to enhance clustering scalability for large document corpora. Their results show that reduced-dimensional embeddings preserve semantic structure while significantly lowering computation time [6].

Kamal et al. explored semantic similarity techniques for academic document classification using embeddings and domain-specific preprocessing. Their work improves clustering accuracy for research papers but requires domain-adaptive preprocessing for best performance [7].

III. PROPOSED SYSTEM

The proposed **AI-Based Automated File Organizer** is designed to intelligently analyze, categorize, and structure files using content-aware machine learning techniques. Unlike traditional file managers that rely solely on filenames and extensions, the system extracts semantic information from documents, images, and scanned files to generate meaningful categories automatically. The solution operates entirely on the local machine, ensuring data privacy while leveraging lightweight LLMs and multimodal processing for real-time organization. The system continuously monitors selected directories, analyzes new files on arrival, and dynamically arranges them into contextually relevant folders without user intervention.

The main objectives of this proposed system are:

- To develop an AI-powered semantic analysis system to automatically create intuitive and scalable folder hierarchies based on file content.
- To design a search module that provides concise file summaries in response to user queries, enabling quick access to relevant file information.
- To design a module to ensure that's, the system operates locally on the user's device, prioritizing data privacy and security.
- To develop a user-friendly desktop application featuring drag-and-drop functionality and a visual folder tree, simplifying file management for users of all skill levels.

Our system supports bulk file upload, allowing users to efficiently drag and drop thousands of file including PDFs, text documents, and images for streamlined processing. It intelligently recognizes common file types to facilitate seamless organization. Leveraging content-based classification, the AI categorizes files into meaningful groups such as "Medical" or "Financial" by analyzing the actual content using Natural Language Processing for text-based files and vision models for images, rather than relying solely on file names or extensions. To enhance usability, the system features dynamic folder creation, automatically generating and updating folders and subfolders based on the classified content, and moving files to their appropriate locations.

© 2025 IJIRCCE | Volume 13, Issue 12, December 2025| DOI:10.15680/IJIRCCE.2025.1312045

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)

IV. METHODOLOGY

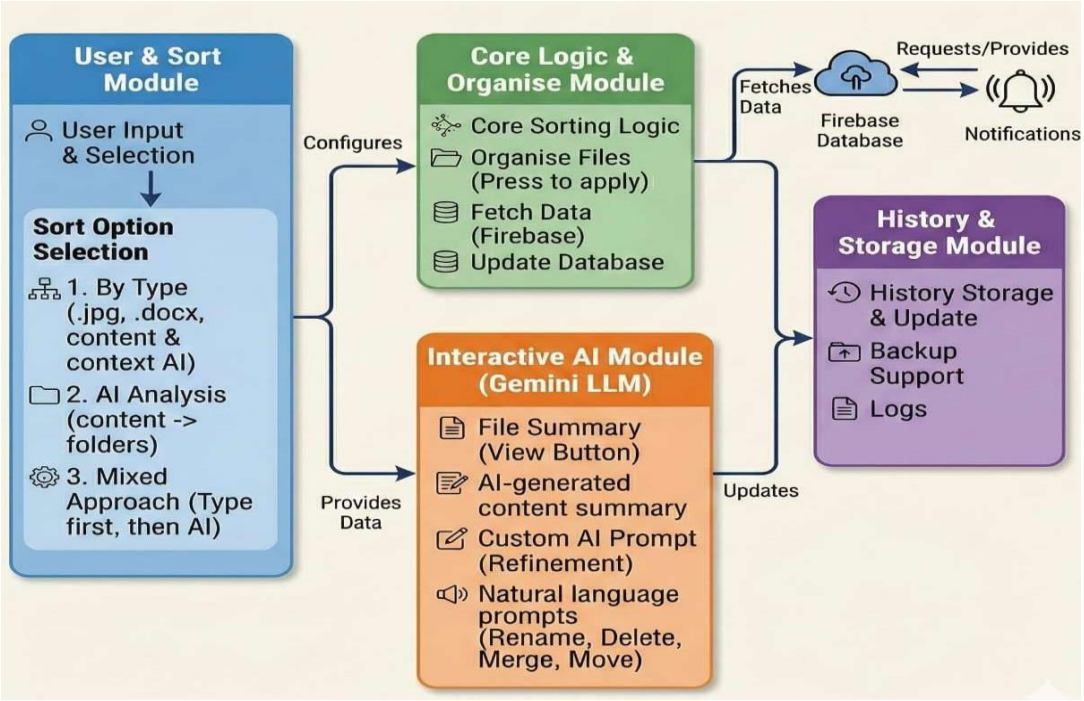


Figure 4.1: Full methodology diagram

The methodology adopted for developing the AI-Based Smart File Organizer follows a structured and modular workflow that ensures efficient file sorting, intelligent categorization, and interactive user control. The overall process integrates user-driven inputs, automated file analysis, and AI-based decision-making to create an intelligent system capable of organizing heterogeneous file types such as PDFs, documents, images, and media files. The figure provided summarizes this workflow through interconnected blocks that capture input selection, sorting logic, analysis pathways, interactive features, and history management. The methodology can be understood in four major stages: User Input Phase, Sorting Option Selection, Core Sorting and AI Analysis, and Interactive and History Management Features.

1. User Input & Selection Phase:

The methodology begins with the user's interaction with the system. The user is first required to select a folder or directory that they wish to organize. This step is essential because the system does not autonomously scan the computer's storage; instead, it relies on user-driven folder input to ensure privacy and focused processing. Once the folder is selected, the system scans the directory and gathers essential details about each file, such as file name, type, extension, size, and metadata. This scanned information forms the basis for further sorting and analysis. The user also participates in subsequent decision-making steps.

2. Sorting Option Selection:

AI-Based Analysis option employs artificial intelligence to analyze the content of files especially PDFs and documents rather than their extensions. The integrated AI model (such as LLaMA running on Ollama) extracts text, interprets meaning, identifies topics, and generates logical categories.

The mixed or hybrid approach combines both the extension-based grouping and AI based semantic analysis. Initially, files are grouped based on their type (e.g., separating documents from images). Then, within the document group, AI performs deeper analysis to further categorize files based on content. This approach balances speed and intelligence by avoiding unnecessary AI processing for non-text files while still providing rich categorization for content-heavy documents. The mixed approach is particularly effective for large directories containing diverse file types.

© 2025 IJIRCCE | Volume 13, Issue 12, December 2025| DOI:10.15680/IJIRCCE.2025.1312045

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)

3. Organize Files Phase:

Once the user selects a sorting option, the system moves to the file organization stage. The "Organize Files" block in the methodology diagram represents the application of the selected sorting logic. At this stage, the system processes all files according to the chosen method. For AI-based sorting, this includes extracting text from PDFs, sending it to the AI model, getting category suggestions, and grouping files accordingly. In this stage, the system creates folder structures that represent the generated categories. For example, a folder named "Research Articles" may be created to store academic materials. The system automatically creates folders, ensures naming consistency, handles duplicates, and moves files to their new destinations. This process requires reliable file system operations, safe handling of file paths, and error recovery mechanisms for problematic files (such as locked or corrupted ones).

4. Core Sorting Logic

The core sorting logic represents the "engine" that powers the entire system. This module implements all sorting decisions, whether they are extension-based, AI-based, or hybrid. It includes several internal functions:

- File scanning and metadata extraction
- PDF text extraction
- AI-based category prediction

5. Interactive Features

The right side of the methodology figure showcases key interactive features designed to enhance flexibility and user experience. The File Summary View improves transparency by letting users click on any file to see its AI-generated summary, detected category, and extracted keywords or content highlights. An Advanced Custom AI Prompt allows users to control folder operations using natural language commands such as "Merge Research and Articles folders," "Rename Finance folder to Billing," or "Move all images to Gallery." Additionally, Natural Language File Operations make the system more intuitive and futuristic than traditional file managers by enabling actions like renaming, deleting, merging, moving folders, and copying files through simple conversational instructions.

6. History Storage & Update: The final block of the methodology highlights the system's history management feature, which stores complete session details after files are organized, allowing users to review previous operations. The saved history includes the folder path, the sorting method used, the categories generated, the number of files sorted, and the timestamp, ensuring transparency and enabling users to track or revisit their past actions.

V. RESULTS

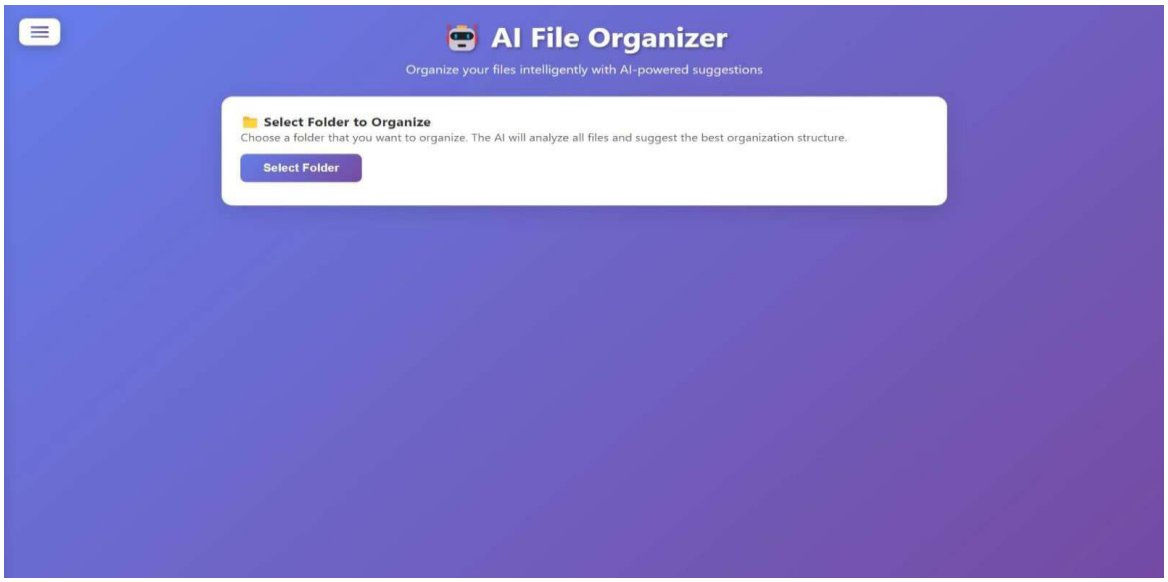


Figure 5.1 : Home Screen of AI File Organizer

© 2025 IJIRCCE | Volume 13, Issue 12, December 2025| DOI:10.15680/IJIRCCE.2025.1312045

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)

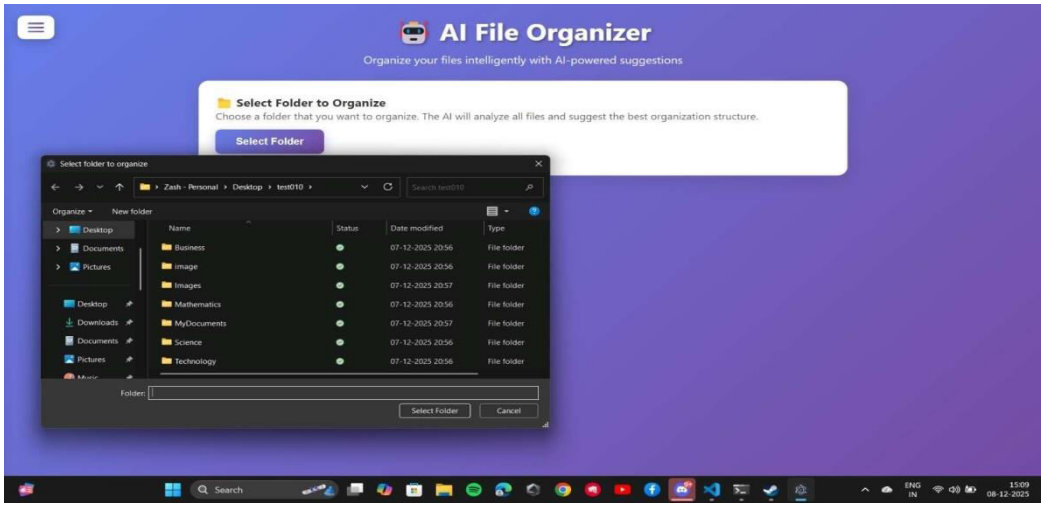


Figure 5.2: Folder Selection Dialog Box for Choosing Target Directory

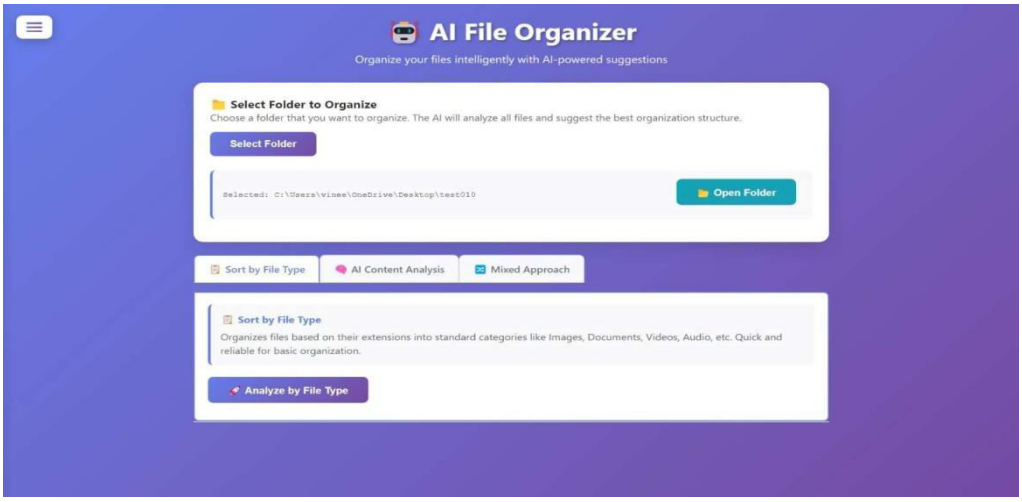


Figure 5.3: Sorting Options Interface

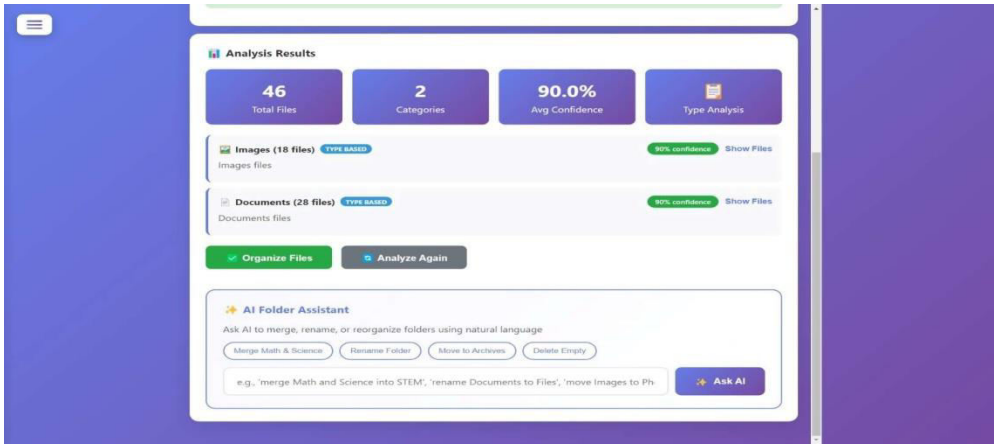


Figure 5.4: Analysis Results with File Counts and Confidence Levels

VI. CONCLUSION AND FUTURE SCOPE

The AI-Based Smart File Organizer successfully demonstrates how artificial intelligence can be integrated into everyday digital tasks to simplify and automate file management. By combining file-type analysis, content-based AI categorization, and a hybrid sorting approach, the system provides a highly accurate and efficient method for organizing large and unstructured file collections. The use of local AI models ensures privacy, speed, and seamless offline functionality while offering intelligent features such as content summaries and natural language commands. Interactive options like folder renaming, merging, and custom prompts make the application adaptable to a variety of user needs.

The AI-Based Smart File Organizer can be extended with several advanced features to further enhance its intelligence, automation, and usability. Future versions can include cloud integration to enable real-time synchronization and multi-device access to organized files. The system can also incorporate advanced OCR for handwritten documents and multilingual support to handle global datasets. Another promising direction is the addition of predictive sorting, where the AI learns user behavior over time and recommends folder structures automatically. Integration with email clients, Google Drive, and corporate data systems can broaden its real-world applicability. Enhanced security features, such as encrypted file movement and access-level control, would make the system suitable for enterprise environments.

REFERENCES

[1] Li, X., Chen, Y., & Wang, L. (2024). Multimodal file organization using integrated text image embeddings for heterogeneous data classification. Journal of Intelligent Information Management, 12(3), 45–58.
[2] Zhang, H., Liu, P., & Zhao, Q. (2023). Transformer-based semantic document clustering using sentence-level embeddings. International Journal of Computer Science and Neural Processing, 18(2), 112–129.
[3] Meta AI Research. (2023). LLaMA: Local large language models for efficient on-device summarization and classification. Meta Research Publications, 1–28.
[4] Yang, J., Xu, M., & Li, T. (2022). CNN-driven image similarity and visual search for scanned multi-format document grouping. Proceedings of the International Conference on Computer Vision Applications, 230–239.
[5] Raza, S., Ahmed, K., & Noor, F. (2022). Content-based file categorization using embedding models and clustering algorithms. Journal of Data Mining and Knowledge Engineering, 10(4), 67–80.
[6] Farzana, N., Rahman, S., & Chowdhury, A. (2021). Dimensionality reduction techniques for large-scale document clustering and visualization. International Journal of Data Science and Analytics, 9(1), 55–70.
[7] Kamal, R., Hussain, M., & Parveen, S. (2020). Semantic similarity approaches for academic document classification and clustering. Journal of Information Retrieval and Intelligent Systems, 7(2), 88–101.



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