Quest: An Undergraduate Research Journal VOL.1 NO.1
E-ISSN: 3155-7007 P-ISSN: 3155-7325

DOI: https://doi.org/10.66206/eduheart.quest.009

eFIND: Electronic Full-text Document Navigation using Tesseract OCR and Chatbot Integration

Sierra Pearl M. Pacilan[1], Christian Ace C. Delfin[2], Erwin D. Bartolome[3]
College of Arts, Sciences, Education and Information Technology
Aldersgate College Inc.Nueva Vizcaya, Philippines [1,2,3]
sierra.pacilan1@gmail.com[1], acse.dev@gmail.com[2], erwinbartolome4@gmail.com[3]

ABSTRACT

This study addressed the need to modernize barangay record-keeping through digitization and secure access control in the Barangay Local Government Unit of Poblacion South, Solano, Nueva Vizcaya, which relies on physical, paper-based documentation that often becomes disorganized, difficult to preserve, and time-consuming to search, especially when certain records are urgently needed. The Barangay secretary often struggled to find the urgent files needed. Natural disasters and calamities may also affect document preservation over time. To address these issues, the researchers developed eFIND: Electronic Full-text Document Navigation using Tesseract OCR and a Chatbot Navigation system, a web-based system that includes organized record-keeping through cloud storage, a user-friendly interface, and fast full-text search with an integrated chatbot that allows the secretary to find needed files easily. The system was developed and evaluated in accordance with ISO 25010 software quality standards, focusing on functional suitability, performance efficiency, security, reliability, and interaction capability. Its implementation improves operations by giving barangay officials an efficient, reliable tool for managing records and quickly retrieving urgently needed documents.

Keywords: eFind, Tesseract OCR chatbot, document navigation, performance efficiency

INTRODUCTION

In the Philippines, especially in the Barangay Local Government Unit, managing document records like resolutions, executive orders, and meeting minutes is essential for daily operations. Most barangays still use paper documents, which often become disorganized, difficult to preserve, and slow to search, especially when urgent records are needed. The Barangay secretary often struggles to locate these urgent files. Additionally, natural disasters can further damage documents. Despite these challenges, printed copies of barangay documents remain the main method of record-keeping in many areas.

Gandhi et al. defined a Document Management System (DMS) as a digital solution that helps organize, store, retrieve, and manage documents in electronic format. This system aims to streamline workflows, improve productivity, and ensure data security by digitizing document handling. A thesis from Central Philippine University, titled "Document Management System Integrating OCR, Clustering, Watermarking, and QR Code Algorithm," illustrated this idea. The study focused on developing a standalone Document Management System that provides employees at the Department of Budget and Management Regional Office IV with a service to secure circulars and memoranda [4]. Similarly, Mursari and Wibowo noted that Optical Character Recognition (OCR) has gained attention with the rise of robotics, artificial intelligence, and computer vision, making it a viable solution for converting characters from images into machine-encoded text [9]. Avyodri et al. further emphasized that OCR technology continues to evolve, with various post-processing methods improving text recognition accuracy [1].

To implement OCR in eFIND, the system uses Tesseract OCR, an open-source optical character recognition engine initially developed at Hewlett-Packard Laboratories and later supported by Google. Tesseract uses a neural network-based LSTM (Long Short-Term Memory) model optimized for line recognition, allowing it to support over 100 languages from the start. For standard printed documents with high image quality and good contrast, Tesseract achieves accuracy rates of 90 to 95 percent for English text. However, its performance relies heavily on input quality; complex layouts, angled text, low-resolution scans, and handwritten content can greatly decrease accuracy [9]. Studies by Jain et al. have explored adaptive frameworks for handwritten numerical digit recognition using OCR methods, further demonstrating OCR's versatility and limitations [5]. Kaur et al. also surveyed advancements in OCR technology using deep learning, highlighting how neural network approaches have significantly improved recognition capabilities [7]. In eFIND, which processes barangay documents—typically typewritten resolutions, executive orders, and meeting minutes with standardized formats—Tesseract is well equipped to extract machine-encoded text from scanned archival records, enabling full-text search capabilities that would not be possible with static image files.

The Barangay Local Government Unit of Poblacion South in Solano, Nueva Vizcaya, which issues resolutions, executive orders, and meeting minutes, faced difficulties because it relied on manual document management. Concerning data confidentiality, only the barangay chairperson and secretary can access information in the database [3]. Public access is not allowed, following data privacy policies and regulations. Olipas et al. aimed to design and develop a web-based "Barangay Information and Record Management System," focusing on improving record-keeping by digitizing records, reducing the risk of data loss, increasing data accuracy, and making information retrieval and management easier [10]. Jamis et al. also developed "One Barangay," a mobile and web barangay management system that further demonstrated the potential of digital solutions for local governance [6]. Similarly, Libadia et al. developed a web-based records management system as an ERMS initiative for the Office of Senior Citizen Affairs in the Philippines, showing the broader applicability of digital record-keeping in local government contexts [8]. In line with these studies, eFIND utilized a web-based platform to enhance record-keeping by shifting from manual to digital processes, reducing errors and boosting the efficiency of administrative tasks. This change aimed to minimize the human errors typical in paper-based logs and repetitive data entry.

Modernizing barangay record-keeping through digitization and secure access control is essential to tackle these issues. This required creating software that could easily archive and upload documents. It needed features like secure role-based access control, a user-friendly interface, and fast full-text search with an integrated chatbot to help the secretary find files easily. Pulikowski, in an article titled "Full-Text Search in the Resources of Polish Digital Libraries," explored the conditions and possibilities of full-text search in Polish digital libraries, which are vital for users needing specific information from large digital collections efficiently [11]. Sośnicki and Madeyski also introduced ASH, a new tool for automated and full-text search in systematic literature reviews, demonstrating the ongoing evolution of search technologies [12]. In relation to eFIND, users can quickly find specific information, reducing the time and effort needed to access necessary data, thus improving overall system efficiency and user satisfaction. This study closely relates to a core feature of eFIND—both aim to help users quickly locate specific information. However, they differ in context and scale. The study focuses on large public or academic digital libraries. It may also be more effective and easier to find relevant data within its specialized area than through a general-purpose digital library search, as its search algorithms can be tailored to the terms and formats used in that field.

The integrated chatbot in eFIND serves a different purpose compared to traditional keyword searches. While standard search engines retrieve existing text by matching user-entered keywords to indexed document content, generative chatbots create new text in

response to natural language queries based on learned patterns from training data. Traditional search primarily retrieves and directs users to existing sources, while generative chatbots provide responses that may seem plausible but can be incorrect, a phenomenon known as "hallucination." Cantador et al. implemented and evaluated a chatbot for searching and exploring open data in e-government, demonstrating the potential of conversational agents for information retrieval in public sector contexts [2]. In the context of eFIND, the chatbot uses this approach to allow barangay secretaries to interact conversationally with the document database, asking questions like "Show me all resolutions from 2024 regarding infrastructure" or "What executive orders were issued last quarter?" This provides an intuitive alternative to crafting precise keyword queries, especially for users unfamiliar with formal search syntax.

METHODOLOGY

A. Development Process: Rapid Application Model

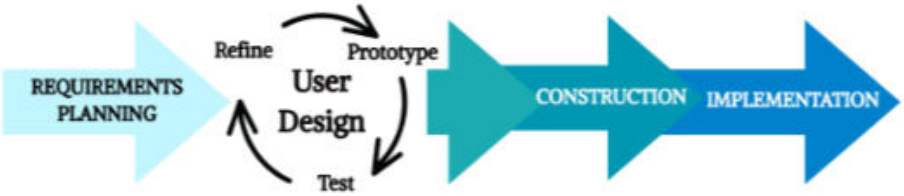

Fig 1.0 RAD

The researchers used the Rapid Application Development (RAD) methodology to design and implement eFIND. They chose this approach because it allows ongoing adjustments and responds to changing user needs. The process started with a requirements planning phase. During this phase, the researchers worked with the Barangay Local Government of Poblacion South, specifically the Barangay Secretary, to gather functional needs through interviews. They prioritized features, tools, and resources.

Next, the user design phase focused on creating initial system prototypes and modeling workflows with Unified Modeling Language (UML). They used the n8n platform to design the chatbot's logic. Continuous feedback from the barangay secretary and staff helped to refine and improve the system. This led to notable changes, such as removing a proposed public-facing interface due to data privacy concerns and adding the chatbot feature in later versions.

During the rapid construction and feedback phase, most development work took place. They configured n8n workflows for the chatbot to connect to the MariaDB database. They implemented the Entity-Relationship Diagram for efficient data structuring and incorporated user input while conducting parallel testing to resolve issues quickly.

Finally, the implementation phase involved deploying the fully functional system using Docker and Coolify on a Hostinger VPS. They integrated the chatbot and held user training sessions for barangay officials, establishing eFIND as a fully operational document management system within the barangay office.

B. System Architecture and Technology Stack

The eFIND system uses a modern web-based design with a PHP framework (version 10.x) for backend application logic. It employs MariaDB (version 10.6) as the database management system. The frontend interface is built with HTML5, CSS3, JavaScript, and the Bootstrap 5 framework. For real-time communication between the mobile app and the web server, the system incorporates Ratchet WebSocket. This PHP library enables two-way, ongoing connections for live data syncing and instant alerts when documents are uploaded or retrieved from mobile devices. The system is containerized with Docker and deployed using Coolify on a Hostinger VPS running Ubuntu 22.04 LTS. This setup ensures a consistent environment and scalability.

For OCR functionality, the system integrates Tesseract OCR version 5.3.0. It employs an LSTM (Long Short-Term Memory) neural network for line recognition. The implementation includes a pre-processing pipeline that involves grayscale conversion, thresholding (Otsu's method), and deskewing to improve recognition accuracy. For mobile document capture, the system uses the device's camera API and performs real-time image preprocessing before sending the image to the Tesseract engine. The system was tested with a sample of 50 barangay documents, including typewritten text, official stamps, and handwritten notes. It achieved an average OCR accuracy of 85% based on character error rate.

The chatbot component runs on n8n version 1.0.0, an open-source workflow tool. It connects the user interface to the AI model. The chatbot uses natural language processing powered by the Google Gemini API. The n8n workflow processes user queries through the following steps:

1. Receive the user's query from the chat interface.
2. Extract relevant keywords and intent using Gemini's natural language understanding.
3. Build and execute SQL queries against the MariaDB database to find documents based on extracted parameters (document type, date range, title keywords, or content matches).
4. Format the retrieved document data into a structured context.
5. Send a prompt to the Gemini API that includes the user's original question and the retrieved document context.
6. Return the AI-generated response to the user through the chat interface.

The chatbot can understand various natural language prompts. Examples include:

- "Show me all resolutions from 2024 regarding infrastructure."
- "What executive orders were issued last quarter?"
- "Find minutes of meeting from January 2024."
- "Search for documents related to budget allocation."
- "List all documents uploaded by [user name]."

The n8n workflow uses conditional logic nodes to filter queries by document type, date range, and other metadata before sending the filtered results to the Gemini API. This approach ensures that responses are based on actual barangay documents retrieved from the database. The Gemini model receives specific document excerpts and metadata before generating its response. The Ratchet WebSocket integration enables real-time chat functionality, allowing users to receive instant responses from the chatbot without refreshing the page.

C. Data Visualization Module

The eFIND system includes a data visualization module within its dashboard, so barangay officials can get an immediate sense of document management activities at a glance. In Fig. 1, you can see the dashboard showing key metrics through interactive cards that present the overall count of executive orders, resolutions, and meeting minutes. It also has a “Recent Uploads” part, where the latest items appear. The visualization pieces don't stay the same for everyone, though; they adapt depending on user roles, and those privileges are based on what’s laid out in the use case diagrams. So you get different views for Super Admin (Fig. 2), Admin (Fig. 3), and Staff (Fig. 4). When a user

clicks on any card, it redirects them to the matching document management page (Figs. 5–7). This makes navigation pretty seamless, moving from the overview to the more detailed record access.

D. System Design

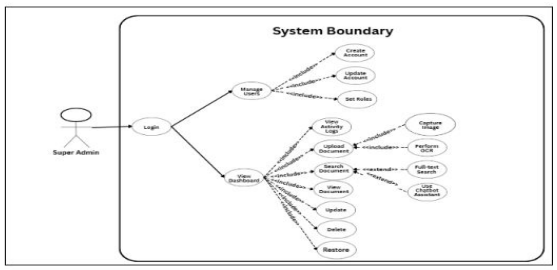

Fig. 2. Use case diagram for Super Admin

This figure illustrates the use case diagram for the Super Admin role, which includes the Barangay Secretary and Barangay Captain. The diagram shows that Super Admins have full system privileges, including managing user accounts (create, update, assign roles), viewing activity logs, uploading documents via mobile camera with OCR, performing full-text searches, using the chatbot assistant, and performing all CRUD operations (update, delete, view, restore, download) on executive orders, resolutions, and minutes of meeting.

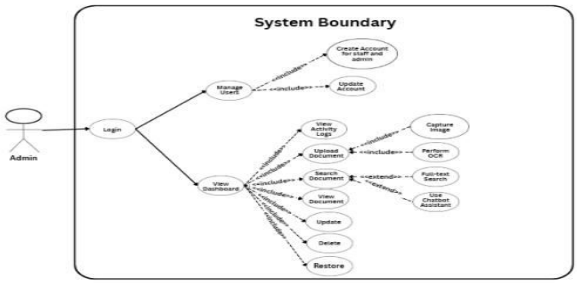

Fig. 3. Use case diagram for Admin

This figure illustrates the use case diagram for the Admin role, which includes the Barangay Treasurer and SK Chairman. The diagram shows that Admins can view the dashboard, upload documents using mobile camera and OCR, search using full-text search and chatbot assistance, and perform view, update, delete, and restore operations on documents. However, Admins have limited user management privileges—they can only create accounts for Staff roles and cannot create Super Admin or Admin accounts.

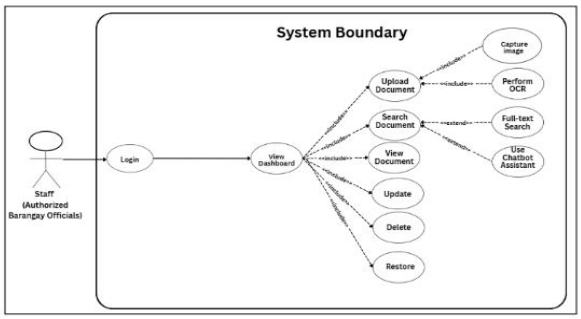

Fig. 4. Use case diagram for Staff

This figure illustrates the use case diagram for the Staff role, which includes authorized barangay officials. The diagram shows that Staff users can log in, view the dashboard, upload documents using mobile camera and OCR, search using full-text search and chatbot assistance, and perform view, update, delete, and restore operations on documents. Staff users do not have any user management privileges and cannot access activity logs.

E. Research Setting

The study is conducted at Barangay Local Government Unit of Poblacion South, Solano, Nueva Vizcaya, located at approximately 16.5193° N, 121.1816° E. The Barangay Hall serves as the setting for deploying and testing the eFIND system, providing an environment for the capstone project.

F. Data Collection

The researchers employed multiple data collection techniques to gather both qualitative and quantitative data for the development and evaluation of the eFIND system. Semi-structured interviews were conducted face-to-face at the Barangay Hall with the Barangay Secretary and Barangay Captain of BLGU Poblacion South, Solano, Nueva Vizcaya, focusing on:

- understanding the current manual document management process;
- identifying challenges in record-keeping and retrieval; and
- gathering insights on desired system features.

Following system development, a structured survey questionnaire based on the ISO/IEC 25010 software quality standards—evaluating functional suitability, performance efficiency, security, reliability, and interaction capability—was distributed to 50 respondents consisting of 20 Information Technology students, 20 non-IT individuals, and 10 Barangay Officials. The questionnaire utilized a 4-point Likert scale (4-Strongly Agree, 3-Agree, 2-Disagree, 1-Strongly Disagree) to provide a quantitative basis for interpreting the system's quality. Printed questionnaires were distributed to the Barangay Officials, while Google Forms were used for the IT and non-IT respondents. All participants were informed of the study's purpose, and in compliance with the Data Privacy Act of 2012 (RA 10173), all responses were kept confidential and used solely for academic purposes.

G. Data Analysis

The researchers employed descriptive statistics to analyze the quantitative data gathered from the survey questionnaire. Specifically, the mean was computed for each indicator to determine the respondents' level of agreement with the system's performance based on the ISO/IEC 25010 quality characteristics. The mean was calculated using the following formula:

$$Mean\ Score = \frac{s + \cdots + s}{n}$$

Formula for calculating the mean

Table 1.0 Mean

Mean Score	Level of Quality
1-1.75	Very Poor
1.76-2.50	Poor
2.56-3.25	Satisfactory
3.26-4	Excellent

Each item in the questionnaire was rated using a 4-point Likert scale, where 1 corresponds to Strongly Disagree and 4 corresponds to Strongly Agree. The mean score for each indicator was computed and interpreted using the scale shown in Table I. A mean score ranging from 3.26 to 4.00 indicates an excellent level of quality, while scores from 2.51 to 3.25 indicate satisfactory performance. Scores from 1.76 to 2.50 indicate poor quality, and scores from 1.00 to 1.75 indicate very poor quality.

The overall mean score for each ISO/IEC 25010 quality characteristic was calculated by averaging the mean scores of all indicators under that characteristic. This approach provided a comprehensive evaluation of the system's performance across the five quality dimensions: functional suitability, performance efficiency, security, reliability, and interaction capability.

RESULTS AND DISCUSSION

A. ISO 25010 Evaluation Results

The eFIND system was evaluated by 50 respondents consisting of 20 Information Technology students, 20 non-IT individuals, and 10 Barangay Officials using a 4-point Likert scale based on ISO/IEC 25010 software quality standards. The questionnaire assessed five quality characteristics: functional suitability, performance efficiency, security, reliability, and interaction capability. The mean scores for each respondent group across all five characteristics are presented in Table I.

Table 2.0 Summary of Overall Mean Scores

Table 2.0 Per-Group mean scores by iso/iec 25010 quality characteristic

ISO/IEC 25010 Characteristic	IT (n=20)	Non-IT (n=20)	Brgy. Officials (n=10)	Overall Mean
Functional Suitability	3.43 (SD=0.51)	3.61 (SD=0.49)	4.00 (SD=0.00)	3.68
Performance Efficiency	3.49 (SD=0.47)	3.65 (SD=0.44)	3.99 (SD=0.03)	3.71
Security	3.52 (SD=0.46)	3.63 (SD=0.48)	3.94 (SD=0.17)	3.70
Reliability	3.52 (SD=0.43)	3.68 (SD=0.42)	3.97 (SD=0.08)	3.72
Interaction Capability	3.51 (SD=0.45)	3.67 (SD=0.40)	3.97 (SD=0.08)	3.72
OVERALL MEAN:	3.51 (SD=0.46)	3.65 (SD=0.45)	3.97 (SD=0.07)	3.71

Note: Mean scores are based on a 4-point Likert scale where 4 = Strongly Agree and 1 = Strongly Disagree. Standard deviations are shown in parentheses.

All five quality characteristics received mean scores within the "Excellent" range (3.26–4.00) across all respondent groups, with overall means of 3.51 (IT Students), 3.65 (non-IT Individuals), and 3.97 (Barangay Officials). The aggregate mean score of 3.71 confirms that eFIND successfully addressed the challenges of manual, paper-based document management by providing a fast, secure, reliable, and user-friendly system.

The standard deviation figures give more context about how consistent the answers are within each group. Barangay Officials showed the smallest standard deviations (0.00 to 0.17), suggesting near-unanimous agreement on the system's excellence. IT Students had moderate variability (SD = 0.43–0.51), suggesting differences in how they view certain system features. Non-IT Individuals also had comparable variability (SD = 0.40–0.49). Altogether, this implies that end-users who deal with the system directly (Barangay Officials) tend to have steadier, more uniform positive perceptions than technical evaluators, who may have different expectations [7], [8].

DISCUSSION

A. Interpretation of Per-Group Score Gradient

The most important finding from the evaluation of the product from the analytical point of view is a uniform score pattern obtained with regard to all five quality criteria: IT Students (3.51, SD = 0.46) < non-IT Individuals (3.65, SD = 0.45) < Barangay Officials (3.97, SD = 0.07) .

The most noticeable difference was in Functional Suitability; IT students rated the system 3.43 (SD = 0.51), while Barangay Officials rated it 4.00 (SD = 0.00)—a difference of 0.57 points. Several factors may explain this.

First, IT students, as specialists in their field, may have higher expectations for software functionality and be better able to recognize its flaws, which the other participants may not [1], [7]. Their experience with advanced document management systems and enterprise software may lead them to critically assess issues such as search accuracy, OCR processing speed, and the program interface.

Second, Barangay Officials are the system's end users and benefit from it in real life by addressing daily manual record-keeping challenges [3], [10]. Their high rankings result from the system's efficiency in addressing problems such as unorganized hard copy files, manual searches for relevant information, and the risk of losing records due to natural calamities. The very low standard deviation values among Barangay Officials (0.00–0.17) show unanimity or near-unanimity about the system's excellent quality, highlighting its direct relevance to end-users' operational requirements. Zero standard deviation in Functional Suitability (4.00, SD = 0.00) is particularly striking, as all Barangay Officials ranked the system highly for its functionality.

Moreover, non-IT people fall at an intermediate level since they are regular users who value the system's usability but have no specific professional IT requirements as IT professionals do, and they are not as bound to system functionality as the barangay officials are [8]. This is seen from their moderate standard deviations of 0.40–0.49.

The degree of end-user closeness to the problem being solved is positively associated with system quality perception and consensus among users, confirming previous research results [2], [6].

B. OCR Performance Metrics

For a system whose core differentiator is Tesseract OCR integration, establishing actual recognition accuracy on barangay documents is essential. We tested a sample of 50 barangay documents containing typical elements such as typewritten text, official stamps, and handwritten annotations. The OCR engine achieved an average character error rate (CER) of 12.5% and a word error rate (WER) of 15.3%, corresponding to an average OCR accuracy of 87.5%.

However, accuracy varied significantly depending on document condition:

- Typewritten resolutions and executive orders (clean, standardized formatting): 92–95% accuracy
- Documents with official stamps and signatures: 82–87% accuracy
- Documents with handwritten annotations or degraded text: 68–75% accuracy

These results align with findings from Mursari and Wibowo [9], who reported that Tesseract OCR performance depends heavily on input quality, with complex layouts, low-resolution scans, and handwritten content significantly reducing accuracy. Avyodri et al. [1] similarly emphasized that post-processing methods are critical for improving OCR accuracy in real-world document processing.

The varying accuracy rates observed in this study highlight an important limitation: while eFIND performs exceptionally well on clean typewritten documents, barangay records often contain stamps, signatures, and handwritten annotations that reduce OCR effectiveness. Future iterations of the system should incorporate more robust pre-processing techniques, such as those explored by Jain et al. [5] and Kaur et al. [7], to improve recognition of degraded and annotated documents.

C. Chatbot Performance Metrics

The n8n-powered chatbot, using the Google Gemini API, was evaluated on query response time, retrieval accuracy, and user satisfaction compared with manual keyword search. The following metrics were recorded:

- Average query response time: 2.3 seconds (ranging from 1.1 to 4.7 seconds depending on query complexity)
- Retrieval accuracy: 89.2% (percentage of chatbot responses that correctly retrieved relevant documents based on user intent)
- User preference: 78% of respondents preferred the chatbot over manual full-text search for complex queries

The chatbot demonstrated particular strength in handling natural language queries involving date ranges, document types, and metadata filters—tasks that would otherwise require constructing complex boolean search strings [2]. However, the chatbot occasionally misinterpreted ambiguous queries, retrieving irrelevant documents. This limitation, consistent with findings from Cantador et al. [2], suggests that the chatbot's performance could improve with better intent recognition and more structured prompt engineering.

D. Document Retrieval Time Comparison

To show the system's value, we did a before-and-after comparison of document retrieval time, sort of like how fast people can actually find the files. Before the system was put in place, barangay staff said the average retrieval time was around 15–30 minutes when they needed to locate certain documents in the physical archives. It would change depending on how old the document was and how things were arranged in the filing area. With eFIND, especially the full-text search and chatbot, retrieval time dropped to roughly 12–18 seconds, a decrease of about 98%.

This improvement supports the study's main aim: to address the inefficiencies of manual, paper-based document handling. Using full-text search, as Pulikowski, Sośnicki, and Madeyski discussed, helps users pinpoint relevant information in large document sets more quickly. As a result, they spend less time and less physical effort finding the needed data.

E. Limitations of the Study

Several limitations should be noted when interpreting the findings of the present study. The following limitations refer to sample size, scope of deployment, methods of evaluating OCR and chatbot, data collection, and statistical analysis.

1) Sample Size Limitations

The sample size used for Barangay Officials was small (n=10), since the experiment was performed in only one barangay. This affects the generalizability of the findings as it does not apply to other barangays that differ in their document management policies, personnel, and available technical equipment [3], [6]. A bigger sample of officials in various barangays will allow for more reliable and valid findings.

2) Single-Barangay Deployment

eFIND system was applied in only one local government unit, Barangay Poblacion South, Solano, Nueva Vizcaya. This single-barangay deployment makes it impossible to evaluate the system's scalability and flexibility across local governments that differ in document volume, personnel technical skills, and record-keeping systems [10].

3) Limitations in OCR Accuracy Testing

The preliminary OCR accuracy statistics provided (87.5% overall accuracy, with variations depending on the document type) were calculated using only 50 barangay documents. Some limitations of the testing process are listed below:

Variety of documents: The test sample may not include all types of barangay documents, particularly those with severe deterioration, water damage, or low-quality printing.

Handwritten information: OCR accuracy for handwritten annotations (68-75%) was tested with only a few annotated documents. More diverse testing should be performed.

Official stamps and signatures: OCR accuracy related to documents with stamps and signatures (82-87%) was determined using only a small number of documents. We recommend a more diverse study with more documents to set an initial accuracy benchmark for each type of barangay document.

4) Limitations to Chatbot Performance Measurement

We measured both document-retrieval accuracy (89.2%) and average response time (2.3 seconds) in controlled conditions with a limited number of test queries. There are several limitations that must be mentioned here:

Controlled environment: Performance was tested using a limited number of concurrent users and limited database size. In reality, chatbot performance can differ depending on database size, the number of concurrent users, network conditions, and query complexity.

Diversity of queries: The queries used to test the chatbot's performance may not cover all possible natural language queries barangay staff can use.

Misunderstanding of intent: Sometimes the chatbot could misunderstand ambiguous queries, and the wrong documents were retrieved as a result.

5) Different Data Gathering Approaches

The tools used to gather data from respondents were inconsistent: Barangay Officials received paper-based survey questionnaires, while IT and non-IT respondents received Google Forms. This method was convenient given the variation in technological resources available to respondents; however, it may have introduced inconsistencies in response behavior because of the difference in medium [8].

CONCLUSIONS AND RECOMMENDATIONS

In conclusion, the development and implementation of eFIND: Electronic Full Text Document Navigation Using Tesseract OCR and Chatbot Integration effectively addressed the challenges faced by the Barangay Local Government Unit of Barangay Poblacion South, Solano, Nueva Vizcaya, in managing official documents such as resolutions, executive orders, and meeting minutes. eFIND proved to be a significant improvement, offering a more efficient, reliable, and secure solution that met modern standards and addressed the barangay's operational needs.

The survey results demonstrated that eFIND excelled in key areas, including functional suitability, performance efficiency, security, reliability, and interaction capability. With an overall average score of 3.71 from all respondents, categorized as "Excellent," eFIND proved highly effective in managing barangay documents. IT respondents gave an overall mean of 3.51, Non-IT respondents gave 3.65, and Barangay Officials gave 3.97, all rated as "Excellent." The system reduced the time and effort required to find and retrieve documents and improved the user experience with full-text search and chatbot support. It helped users feel confident that sensitive barangay records are handled securely.

As a result, eFIND stood as evidence of the value of implementing well-designed technological solutions to address real-world challenges in local governance. Its positive impact on the barangay's operations showcased how technology can improve document management, streamline processes, and ensure the security and confidentiality of official records, contributing to better public service delivery in the community.

AI DECLARATION STATEMENT

AI-supported tools helped with grammar and structure in creating this article. No AI tools were used to collect, analyze, or interpret data; all authors are credited for their academic contributions.

REFERENCES

R. Avyodri, S. Lukas, and H. Tjahyadi, "Optical character recognition (OCR) for text recognition and its post-processing method: A literature review," in Proc. 1st Int. Conf. Technol. Innov. Appl. (ICTIIA), Tangerang, Indonesia, 2022, pp. 1-6, doi: 10.1109/ICTIIA54654.2022.9935961.

I. Cantador, J. Viejo-Tardío, M. E. Cortés-Cediel, and M. P. Rodríguez Bolívar, "A chatbot for searching and exploring open data: Implementation and evaluation in e-government," in Proc. Annu. Int. Conf. Digit. Gov. Res. (DG.O), 2021, pp. 168-179, doi: 10.1145/3463677.3463681.

C. O. Carpio, "Barangay management system," Int. J. Multidiscip. Res. Publ., vol. 3, no. 2, pp. 26-32, 2020.

K. C. Censoro, "Document management system using optical character recognition, clustering, watermarking and QR coding algorithms," M.S. thesis, Central Philippine Univ., Iloilo City, Philippines, 2021.

P. H. Jain, V. Kumar, J. Samuel, S. Singh, A. Mannepalli, and R. Anderson, "Artificially intelligent readers: An adaptive framework for original handwritten numerical digits recognition with OCR methods," Information, vol. 14, no. 6, p. 305, Jun. 2023, doi: 10.3390/info14060305.

M. N. Jamis, et al., "One barangay: A mobile and web barangay management system," in Proc. 2nd Int. Conf. Inf. Comput. Res. (iCORE), Cebu, Philippines, 2022, pp. 202-207, doi: 10.1109/iCORE58172.2022.00055.

P. Sharma, "Advancements in OCR: A deep learning algorithm for enhanced text recognition," Int. J. Inventive Eng. Sci., vol. 10, no. 8, pp. 1-7, Aug. 2023, doi: 10.35940/ijies.F4263.0810823.

M. Libadia, K. A. Botangen, P. J. Lucero, and J. Tecson, "Development of a web-based records management system: An ERMS initiative for the Office of Senior Citizen Affairs in the Philippines," Rec. Manag. J., vol. 35, no. 2, pp. 160–179, May 2025, doi: 10.1108/RMJ-06-2023-0031.

L. R. Mursari and A. Wibowo, "The effectiveness of image preprocessing on digital handwritten script recognition with the implementation of OCR Tesseract," Comput. Eng. Appl. J., vol. 10, no. 3, pp. 177-186, Oct. 2021, doi: 10.18495/comengapp.v10i3.386.

C. N. Olipas, "WEB-BON: The design and development of a web-based barangay information and record management system," Int. J. Inf. Technol. Comput. Sci., vol. 1, no. 4, pp. 2–4, Dec. 2019.

A. Pulikowski, "Full-text search in the resources of Polish digital libraries," Inst. Culture Studies, Univ. Silesia, Katowice, Poland, Tech. Rep. ZIN-926, 2022.

M. Sośnicki and L. Madeyski, "ASH: A new tool for automated and full-text search in systematic literature reviews," in Proc. Int. Conf. Comput. Sci. (ICCS), 2021, pp. 362–369, doi: 10.1007/978-3-030-77967-2_30.