

Chapter 4 - RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the empirical results of the evaluation of the AI-enhanced document governance system. The findings directly address Research Question 4 (RQ4) by systematically assessing the prototype against the success criteria derived from the methodological framework established in Chapter 3.

This chapter is organized into two main parts to provide the necessary context for the evaluation. First, it presents the **developed artifact**—the final proof of concept (PoC) prototype—showcasing its key features and user interface. Second, it will detail the results of the formal evaluation, which are structured into three sections.

1. **Expert Validation:** Assessing the system's architectural and governance soundness.
2. **User Testing:** This section measures the prototype's usability and effectiveness with SGU staff.
3. **System and Governance Validation:** Testing the system's compliance with non-negotiable security and reliability rules.

The chapter concludes by synthesizing these results and discussing their implications in relation to the research questions and hypotheses.

4.2 The Developed Artifact: System Overview

This section describes the artifact—the **AI-enhanced Document Governance System prototype**—that was developed as the core output of this research. It details the system's architecture, its key functionalities, and its initial technical performance, providing essential context for the subsequent evaluation sections.

Although it was designed to replicate the Google Workspace for Education Plus (GWS-E+) ecosystem of Swiss German University (SGU), the prototype was developed and tested on the researcher's personal GWS-E+ platform. This strategic decision was made to circumvent the significant procedural complexities, potential costs, and administrative delays associated with directly integrating into SGU's live institutional environment for a research-level proof of concept (PoC). The prototype's architecture and configuration are functionally identical to how it would operate within the SGU environment. This ensures that the evaluation results are directly transferable and relevant.

4.2.1 System Architecture

The implemented prototype is built on a multi-tiered, layered architecture that provides clear separation of concerns to ensure modularity, scalability, and maintainability. The system is structured into five distinct layers, as detailed in Table 4.1 and illustrated in Figure 4.1, ensures the system is scalable and easy to maintain by creating a clear separation of concerns between the user interface, the core business logic, the data storage, the AI intelligence, and the notification systems.

Table 4.1 System Architecture Layers

Layer	Technology	Function
User Interface	Google AppSheet	Provides the web-based front-end for staff to upload documents, perform searches, and interact with the system.
Automation & Integration	Google Apps Script	Serves as the central business logic, orchestrating workflows between services and triggering AI functions.
Database	- Google Drive - Google Sheets	Drive: Secure, centralized repository for document files. Sheets: Structured database for all document metadata (PoC).
AI Services	- Vision AI - Vertex AI - Gemini AI	Provides the system's intelligence, including OCR, metadata classification, and semantic search.
Notification	Google Calendar	Manages and sends automated alerts and reminders for contract deadlines.



Figure 4.1 System Architecture Layers

4.2.2 Key Features and User Workflow

The prototype encompasses several interconnected features designed to address the identified challenges in SGU's document governance.

4.2.2.1 Automated Document Upload and Classification:

The system employs a robust, multi-step workflow for automated document upload and classification. This process, initiated through a user-friendly interface and orchestrated by **Google Apps Script**, ensures that documents are accurately categorized and organized, even when a user's initial input is incorrect.

The workflow, as illustrated in Figure 4.2, proceeds as follows:

1. **User Authentication and Role-Based Access Control (RBAC):** Users must log in to the **AppSheet UI** using their official *sgu.ac.id* account. **Apps Script** then authenticates the user and checks their role to determine their department and grant them access to specific Google Drive folders, ensuring that they can only upload documents to their authorized location.
2. **User Input and Temporary Data Storage:** The user uploads a document, providing an initial title and metadata. **Apps Script** stores the document in temporary **Google Drive** folder and creates a metadata record in **Google Sheets**.
3. **Intelligent Text Extraction:** Apps Script checks the document's file type. If it is a scanned document (e.g., a PDF), it triggers **Google Cloud Vision AI** to perform Optical Character Recognition (OCR), extracting all text from the image.
4. **AI-Powered Classification and Correction:** The extracted text and the user-provided metadata are sent to a custom machine learning model in **Google Cloud Vertex AI**. This model performs two critical functions:
 - **Classification:** It intelligently classifies the document into the correct category (e.g., "MOU," "Contract," or "SOP").
 - **Metadata Correction:** It extracts key entities from the document and corrects any discrepancies in the user's initial metadata. For example, it will override an incorrect document type or a wrong renewal date.
5. **Final Organization and Notifications:** The corrected metadata is then sent back to Apps Script. It uses this information to finalize the process by:

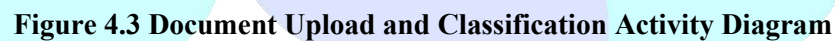
- Updating the metadata record in **Google Sheets**.
- Moving the document from the temporary folder to the correct, designated **Google Drive** folder based on its verified classification and the user's role.
- Creating a reminder in **Google Calendar** and sending an email notification for time-sensitive documents like contracts.

This integrated and self-correcting workflow ensures that documents are accurately processed and organized with minimal manual intervention, while also maintaining a reliable and auditable record.

This structured input process, enhanced by the AI's ability to correct user errors, ensures highly accurate classification and improves document organization and proactive lifecycle management. The AppSheet Upload form is shown in figure 4.2 and the user upload activity is illustrated in the figure 4.3 below.

The figure displays three panels of the AppSheet Document Upload Interface. The left panel shows a list of documents with details like 'doc SOP', 'doc SK aslinya SOP', 'new folder', 'SK_Board_Pengangkatan Rektor 23-27', 'Kotrak_Board_Sewa Gedung Prominence_234-26', and 'sasdc'. The middle panel is a form for adding a new document, with fields for Title, Description, Doc Type (dropdown), and Document (file upload). The right panel is a form for adding a new document, with fields for Title, Description, Doc Type (dropdown with 'Sewa Gedung' selected), Expiry Date (calendar icon), and Email pihak ke-2.

Figure 4.2 The AppSheet Document Upload Interface



The system offers a dual-faceted approach to document retrieval, providing both traditional keyword search and advanced semantic search to accommodate various user needs. This workflow, as illustrated in Figure 4.5, is orchestrated by **Google Apps Script** and is integrated seamlessly within the **AppSheet UI**.

- For a traditional keyword query, **Apps Script** performs a direct search on the structured metadata stored in **Google Sheets**.
- This method is highly efficient for retrieving documents based on exact matches for titles, categories, or other predefined metadata fields.

Sari Nusivera

- For a more complex, conversational query (e.g., "Find all documents about financial policies from last year"), **Apps Script** sends the request to the **Google Gemini API**.
- The Gemini API then performs a **semantic search** across the document metadata and content, understanding the intent behind the user's query rather than just looking for keywords.
- This AI-powered approach provides a highly relevant search result set and can even generate a summary or a direct answer from the document content.

Regardless of the search method used, the final output is presented to the user through the AppSheet UI as a list of search results. Users can then view or download the corresponding document by clicking on a provided symbol or a direct **Google Drive** link, which is generated dynamically by Apps Script. This dual-search functionality ensures that users can quickly and effectively find the information they need while maintaining a secure and controlled access to the document repository.

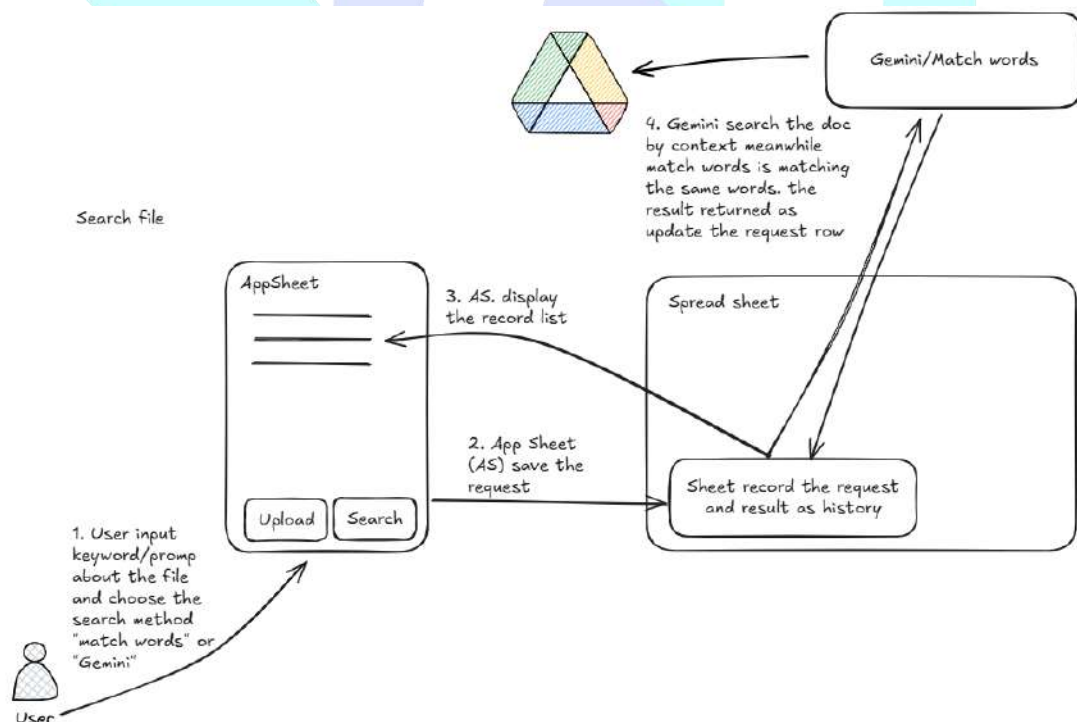


Figure 4.4 Search & Retrieve Activity Diagram

Q&A AI Assistant: The prototype incorporates a conversational AI assistant, leveraging Gemini 2.5 implementation for document Q&A. Users can interact with the assistant to ask questions about document content in natural language. The evaluation of response accuracy and relevance focused on the assistant's ability to provide correct and contextually appropriate answers derived from the indexed documents. Insights were gathered on user interaction patterns and effectiveness to understand how users engaged with the AI and its utility in quickly finding information.

The figure below illustrates the search interface and a sample question-and-answer interaction using Gemini. Users can ask questions in Bahasa Indonesia using simple and conversational language to retrieve accurate information from the relevant institutional documents. This natural language requirement enhances user accessibility and allows even non-technical staff to efficiently extract targeted insights without needing advanced query skills.

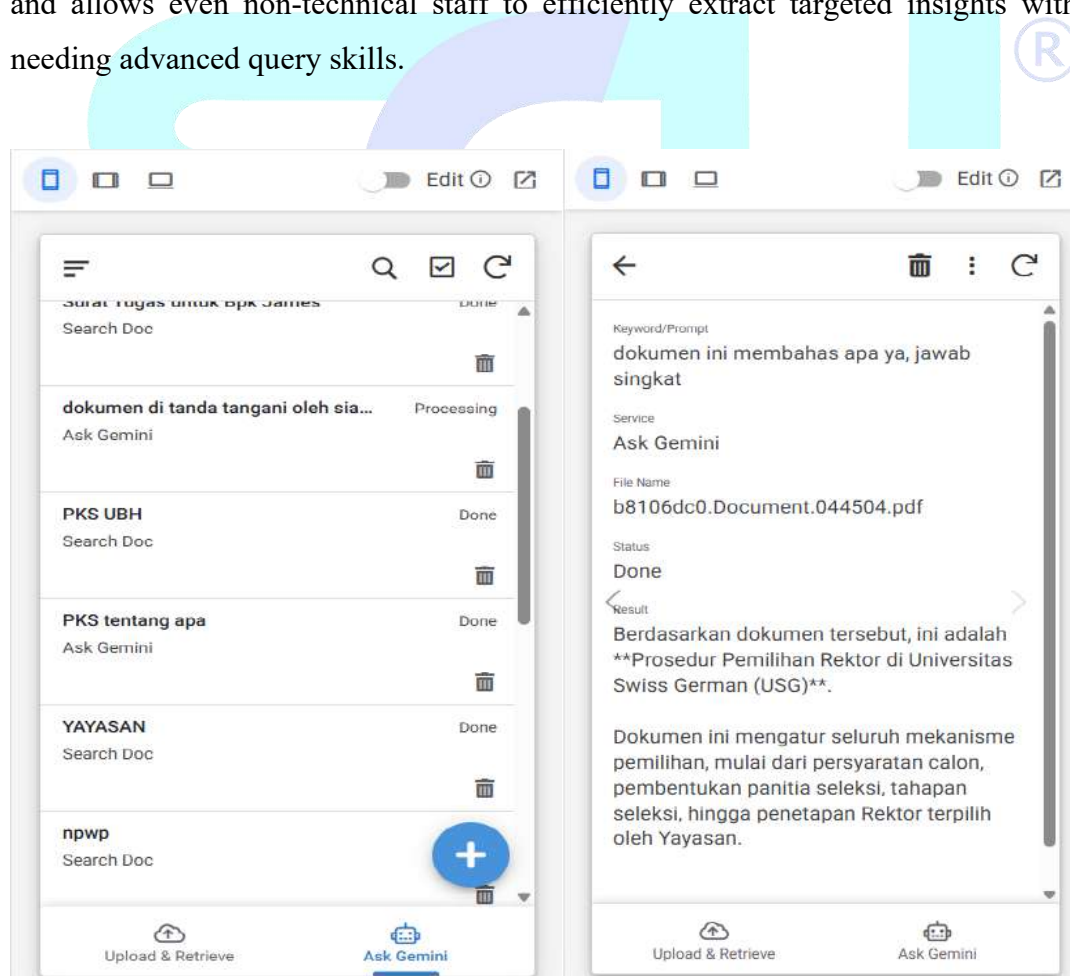


Figure 4.5 Semantic Search and Q&A Assistant Interface

4.2.2.3 Automated Contract Deadline Management: A critical component for risk management, the prototype incorporates an automated contract expiration tracking system. This system identifies contract expiry dates from extracted metadata and automatically generates alerts. Google Calendar integration for notifications ensures that relevant stakeholders receive timely reminders of upcoming contract deadlines. The assessment of the reliability of automated alerts centred on ensuring notifications were accurate and delivered promptly, with the objective of reducing the risk of missed renewals.

The figure below shows the automated email reminder sent to a user upon uploading a document categorized as a contract with an expiry date set for September 25, 2025.

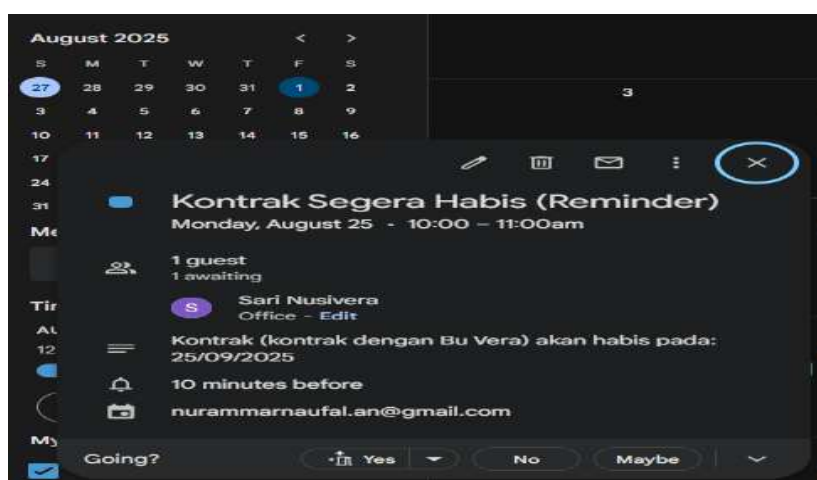
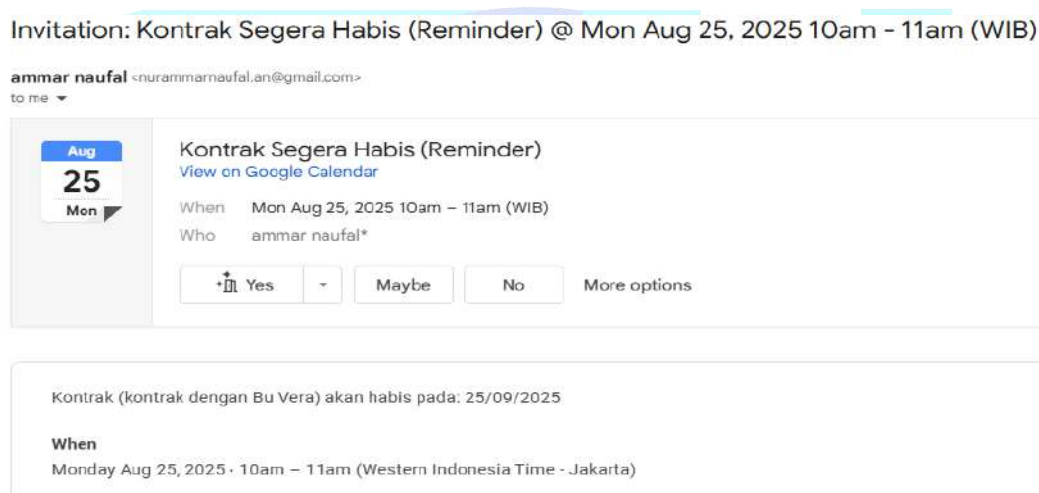


Figure 4.6 Automated Contract Alert (Email and Calendar)

4.2.2.4 Role-Based Access Control and Security: Access control is implemented within the existing Google Workspace environment, leveraging its robust security features. This includes restricting upload rights by department, providing view-only access across the board, and limited cross-department retrieval based on organizational role, as per SGU's initial requirements. A preliminary compliance assessment with COBIT principles regarding access control was considered. This assessment aligned with aspects of APO13 Manage Security and DSS05 Manage Security Services. This was done to ensure foundational security measures and data protection mechanisms were in place. These mechanisms emphasized confidentiality, accountability, and availability.

Role-based access control is assigned to each user account that logs into the AppSheet application. Access to the system is strictly limited to users with a valid @sgu.ac.id email address, which is part of SGU's Google Workspace. Each account must be registered and configured by the system administrator. The administrator assigns a specific access role based on the user's department within SGU's organizational structure. This access role determines two key permissions:

- **Upload Access** – Users must upload documents into folders that belong to their assigned department.
- **Retrieval Access Policy:** Users can only view and download documents from their own department's storage. They cannot access or retrieve documents stored in other departments' folders.

Sensitive or confidential folders are subject to additional restrictions, even within the same department. Only authorized staff can view or manage these documents based on internal clearance levels. Shared folders between departments allow for collaborative work by granting specific access to selected users outside the primary department. Finally, high-level management users will be granted full access to all departmental folders, depending on their roles and responsibilities. All access control and folder-sharing configurations must strictly follow SGU's internal data governance policies and guidelines.

The figure below illustrates the role-based access control mechanism.

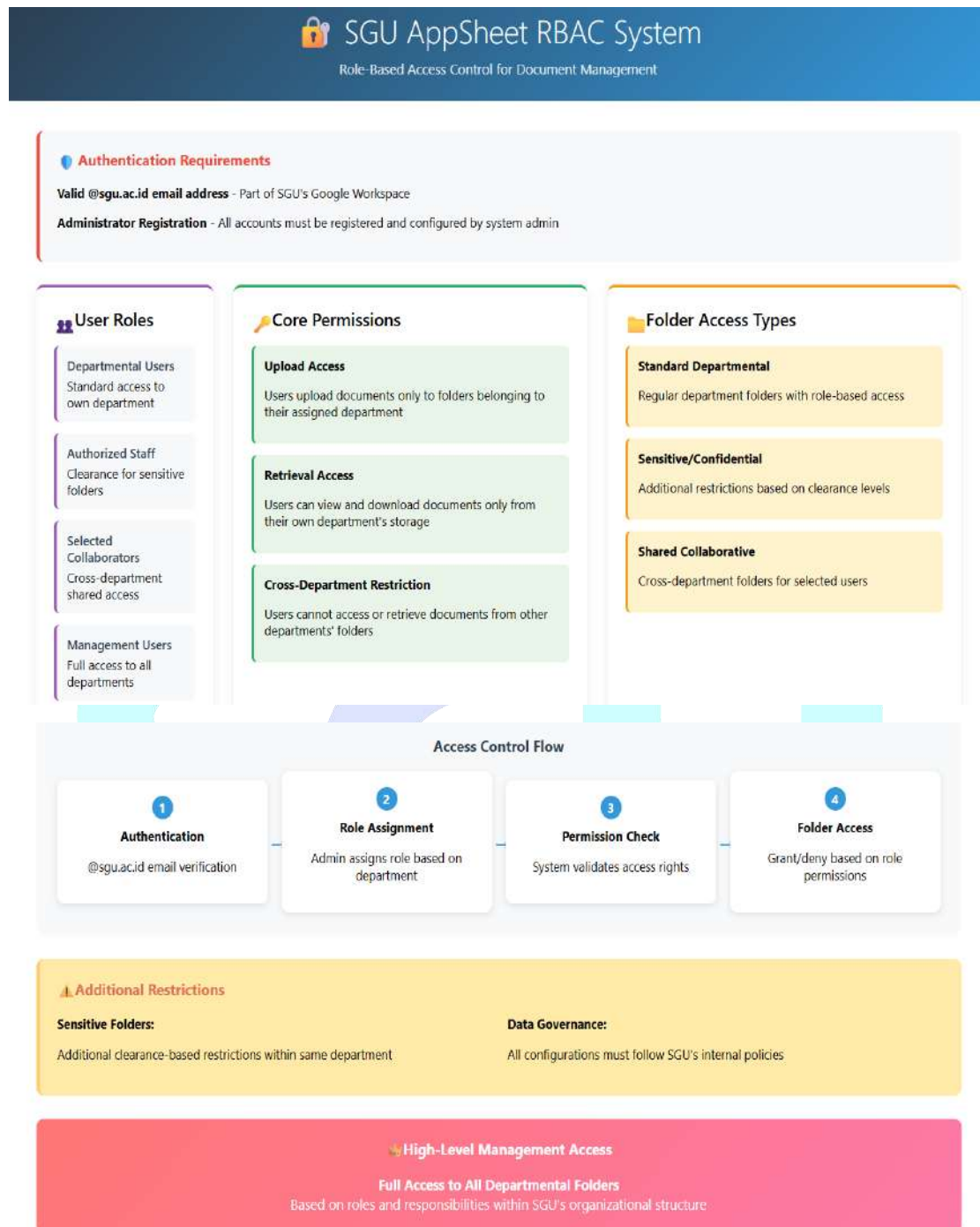


Figure 4.7 Role-Based Access Control Structure

4.2.3 Initial Technical Performance

Initial technical performance indicators were monitored during the prototype's development and internal testing phases to assess its operational viability.

- **System response times and processing speeds** were measured for key functions such as document upload, classification, metadata extraction, and search queries, with the aim of achieving efficient user interaction.
- **The accuracy rates of AI models were measured for OCR**, document classification, and metadata extraction, providing a foundation for understanding the effectiveness of the AI in automating essential tasks.
- **Scalability considerations and initial limitations observed** primarily centred around the PoC's reliance on Google Sheets as the metadata database. This posed inherent constraints for managing large volumes of complex data and concurrent user access in a university-wide deployment.
- **Assessment of integration stability and reliability** across various Google services (Drive, AppSheet, Vision AI, Vertex AI, Gemini AI, Calendar) ensured seamless data flow and functionality.

4.3 Expert Validation Results

This section presents the expert panel's findings from their assessment of the prototype. The evaluation aimed to validate the system's architectural soundness and compliance with governance standards by directly testing hypotheses related to the design's feasibility and adherence to guiding frameworks, such as TOGAF, ISO, and COBIT.

4.3.1 Expert Panel and Methodology

To ensure a multidimensional assessment of the prototype, **a panel of five experts** was assembled. The panel included professionals with diverse and relevant backgrounds, such as senior leaders with expertise in **IT governance, business strategy, human resources, and quality assurance**. This composition ensured that the evaluation would be considered from strategic, operational, and compliance standpoints. The panel's credibility was further underscored by their significant experience; the majority had **over 10 years of experience in higher education** and were **familiar with the Google Workspace platform**. Validation was conducted via a focus group discussion (FGD)

and a structured questionnaire. (Detailed demographic data for the expert panel is provided in Appendix C).

4.3.2 Quantitative Findings: Architectural and Compliance Ratings

The experts evaluated the prototype based on the core architectural and compliance criteria outlined in Chapter 3. The panel's assessment was highly positive, confirming the design of the prototype and its alignment with international standards.

Figure 4.8 shows the experts' overall evaluation of the system's ability to meet its foundational goals. This figure serves as primary evidence of the **TOGAF architectural fit**, showing that the experts agreed the design was coherent and fit for purpose.

In your view, are the stated research objectives (problem-solving focus, architecture design, and system prototyping) appropriate and clearly aligned... the institutional context and research questions?
4 responses

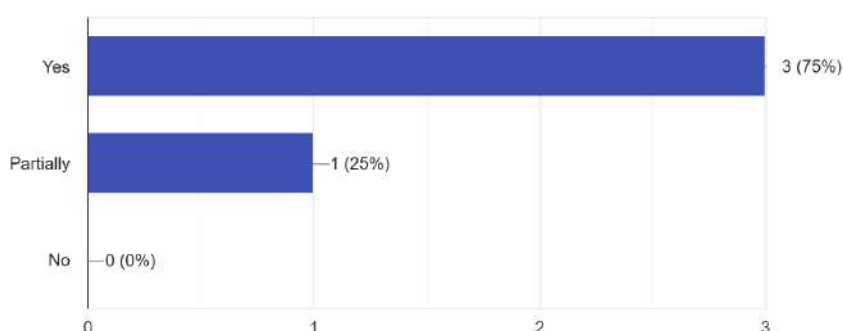


Figure 4.8 Expert Evaluation of Research Objectives Alignment (TOGAF)

Furthermore, the experts validated the system's compliance with key governance standards. The prototype's adherence to **COBIT** principles (Figure 4.9), **ISO 15489** for records management (Figure 4.10), and **ISO 23081** for metadata (Figure 4.11) was strongly affirmed by the panel.

Does the proposed system sufficiently address minimum compliance requirements in document and contract governance — specifically regarding: (a... (c) notification for contract renewal deadlines?

4 responses

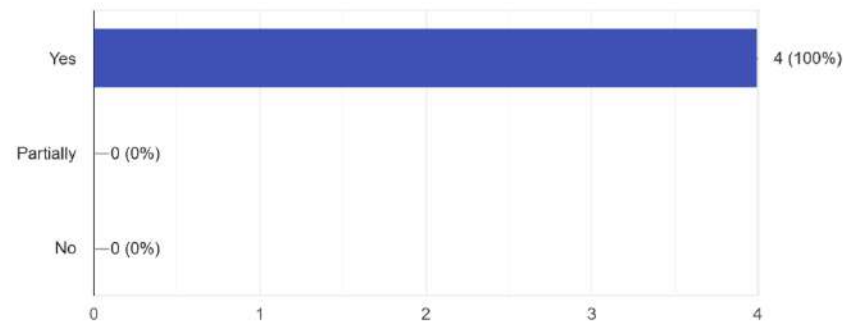


Figure 4.9 Expert Compliance Assessment

Classification (Vertex AI)

4 responses

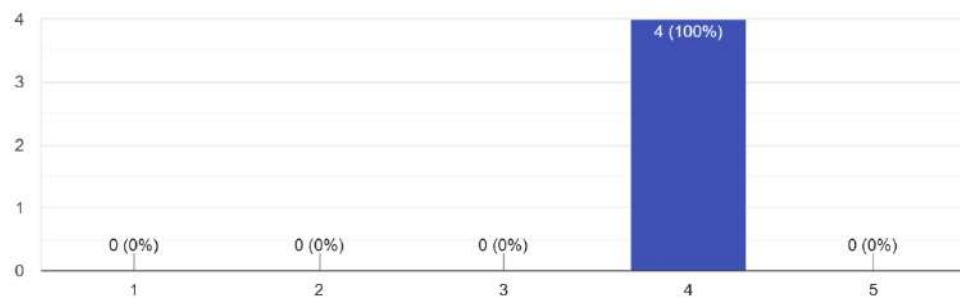


Figure 4.10 ISO 15489 records management compliance

Metadata Governance

4 responses

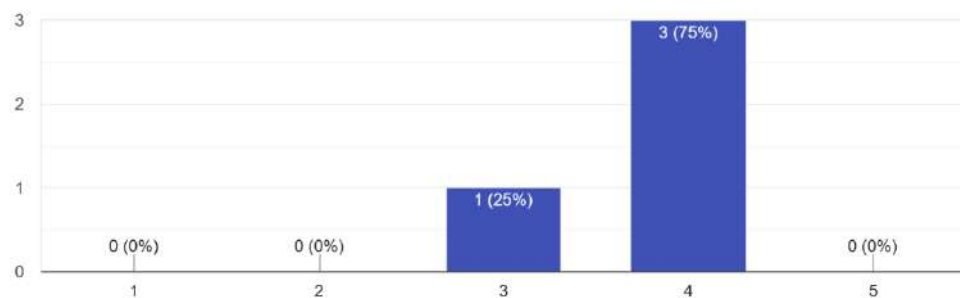


Figure 4.11 ISO 23081 metadata standards compliance

A key component of the architecture is the role-based access control (RBAC) model. As illustrated in Figure 4.12, experts determined that the proposed RBAC structure is technically and operationally sound for SGU's specific requirements.

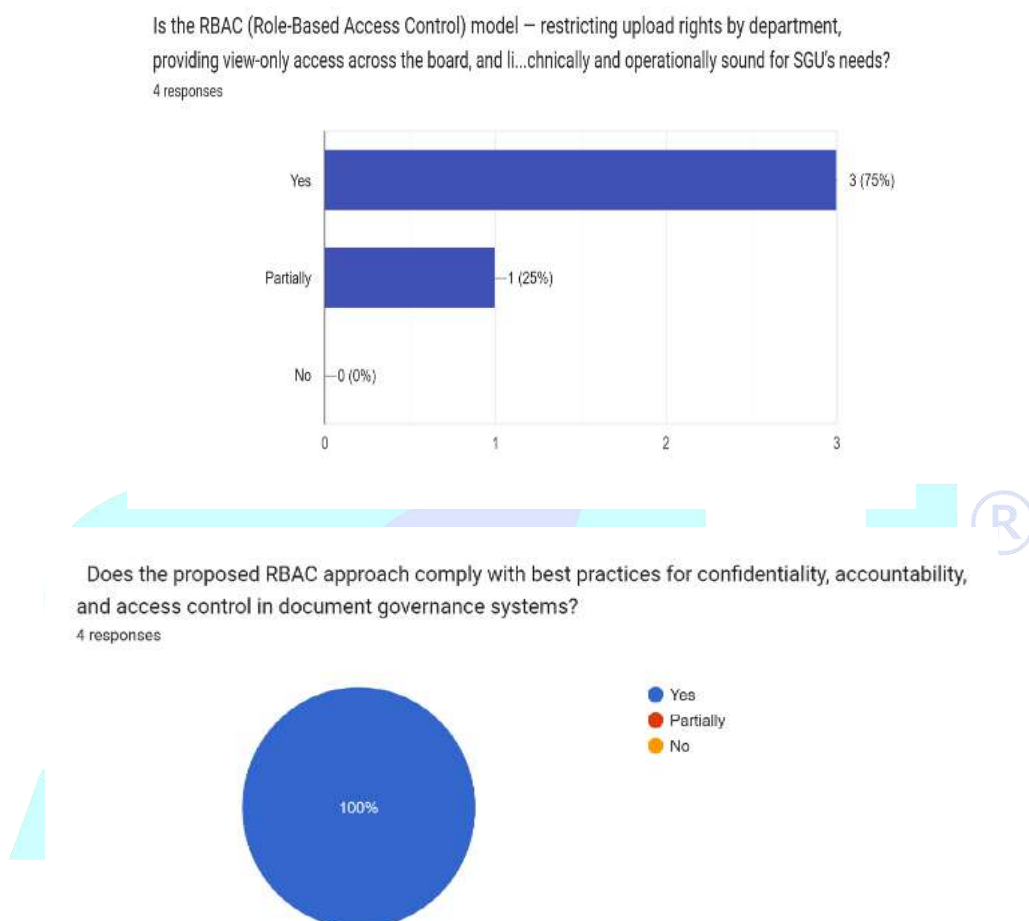


Figure 4.12 Expert Assessment of RBAC Model Soundness

In addition to expert ratings, direct system validation was conducted on two key performance metrics.

- **Metadata completeness:** To test compliance with ISO 23081, the system's metadata repository was audited. As Figure 4.13 shows, two early development uploads failed validation before the rules were finalized. However, the system achieved 100% metadata completeness in its final, steady-state configuration.

1	ID	Description	Department	Doc Type	Expiry Date	Date	Document	Validate
2	7862ea40		HR	SOP		2025-07-14 15:45:41	HR/SOP/7862ea40.Document.0847	TRUE
3	7faf8684		HR	SOP		2025-07-14 15:48:37	HR/SOP/7faf8684.Document.084851	TRUE
4	58b5ddfb		HR	MOU-PKS	07/18/2026	2025-07-14 15:52:12	HR/MOU-PKS/58b5ddfb.Document.0	TRUE
5	63a46c50		HR	SK		2025-07-18 10:21:06	HR/SK/63a46c50.Document.032151	TRUE
6	c748833d		HR	Sewa Gedung	3/25/2026	2025-07-18 10:37:11	HR/Sewa Gedung/c748833d.Docum	TRUE
7	dde9bd55		Finance	Kontrak Vendor		2025-07-18 11:22:10	Finance/Kontrak Vendor/dde9bd55.I	TRUE
8	b8d5ba00		Finance	Sewa Ruko-Gedung	3/18/2026	2025-07-18 11:51:07	Finance/Sewa Ruko-Gedung/b8d5ba	TRUE
9	09af9fcf		HR	SOP		2025-07-22 06:48:29	HR/SOP/09af9fcf.Document.014945	TRUE
10	ec331ca4		HR	SK		2025-07-22 11:33:11	HR/SK/ec331ca4.Document.043521	TRUE
11	5154431b		Finance	Kontrak Vendor	07/23/2025	2025-07-22 11:33:11	Finance/Kontrak Vendor/5154431b.I	TRUE
12	9e85c887		HR	SK		2025-07-22 11:33:00	HR/SK/9e85c887.Document.04874	TRUE
13	b56e072f		Finance	Laporan Keuangan		2025-07-22 11:38:05	Finance/Laporan Keuangan/b56e07	TRUE
14	46f2e436		Finance	Laporan Keuangan		2025-07-22 11:33:24	Finance/Laporan Keuangan/46f2e4	FALSE
15	8ceb07d7		Finance	Kontrak Vendor	08/08/2024	2025-07-22 11:33:15	Finance/Kontrak Vendor/8ceb07d7.I	TRUE
16	5554fe01		HR	MOU-PKS	7/3/2027	2025-07-22 11:37:04	HR/MOU-PKS/5554fe01.Document.1	TRUE
17	f40a5ab2		Finance	Laporan Keuangan		2025-07-22 11:40:44	Finance/Laporan Keuangan/f40a5ab	TRUE
18	618ba7ee		Finance	Kontrak Vendor	08/08/2024	2025-07-22 11:42:35	Finance/Kontrak Vendor/618ba7ee.I	TRUE
19	b8106dc0		HR	SOP		2025-07-22 11:41:42	HR/SOP/b8106dc0.Document.0445	TRUE
20	c4a49db9		Finance	Bukti Transfer		2025-07-22 11:46:18	Finance/Bukti Transfer/c4a49db9.Dc	TRUE
21	29e985a8		HR	SK		2025-07-22 11:47:00	HR/SK/29e985a8.Document.044821	TRUE
22	2224c065		HR	LDP		2025-07-22 11:48:16	HR/LDP/2224c065.Document.04481	TRUE

Figure 4.13 Metadata validation view (PII redacted).

The Validate flag checks required fields and controlled-vocabulary conformance (ISO 23081). Result: 37/39 TRUE (94.9%) overall; 2 early FALSE during development; 100% steady-state after rule hardening.

- **AI Classification Accuracy:** The custom Vertex AI model's accuracy was tested against a validation set. The model achieved a classification accuracy of 94.7%, surpassing the success criterion of 90%, as shown in Figure 4.14.

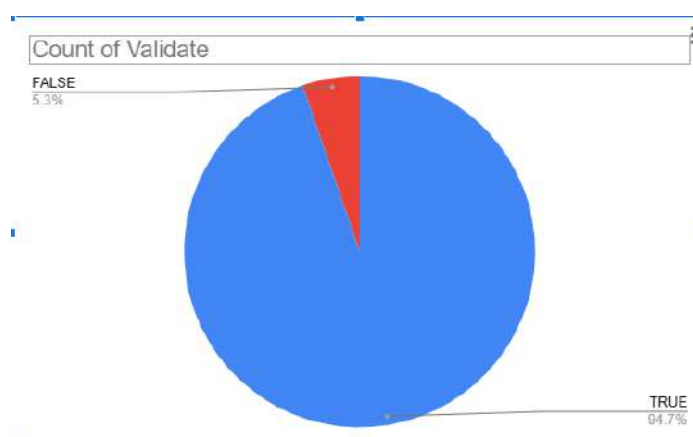


Figure 4.14 Document Classification Accuracy

This comprehensive set of quantitative data confirms the soundness of the system's design and the performance of its core AI and governance functions at or above the required thresholds.

4.3.3 Qualitative Findings: Thematic Analysis

A thematic analysis of the experts' discussions revealed the system's key strengths and important considerations for future development.

Experts have identified the system's capacity for centralized standardization and enhanced accessibility as notable strengths, as visually represented in Figure 4.15: Expert Qualitative Feedback Word Cloud (Key Strengths).



Figure 4.15 Expert Qualitative – Identified Strengths (Word Cloud)

There are several areas that require attention to ensure optimal functionality and compliance. These include the need for mechanisms that are *"access-request friendly,"* addressing system limitations regarding *"decision-making purposes,"* and considering policy implications for user adoption. Figure 4.16: Expert Qualitative Feedback Word Cloud (Areas for Improvement/Suggestions) provides the collective expert insights on areas for improvement and future suggestions. Additionally, experts raised concerns about data safety, the false usage of documents, and the importance of addressing change management for broader adoption beyond the PoC. Their policy recommendations and identified risks are comprehensively mapped in Figure 4.17: Expert Qualitative Feedback - Policy Recommendations (Thematic Map) and Figure 4.18: Expert Qualitative Feedback - Identified Risks/Concerns (Thematic Map), respectively.

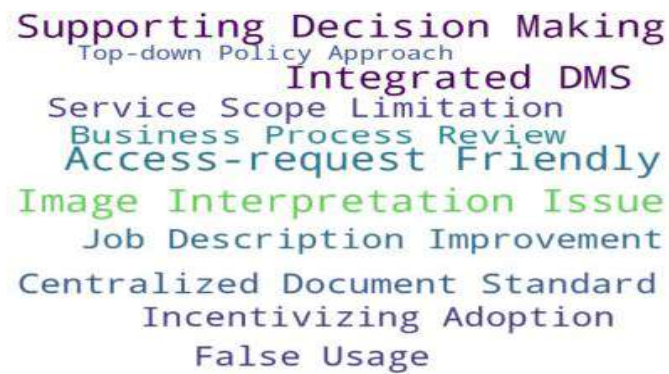


Figure 4.16 Expert Qualitative-Areas for Improvement (Word Cloud)

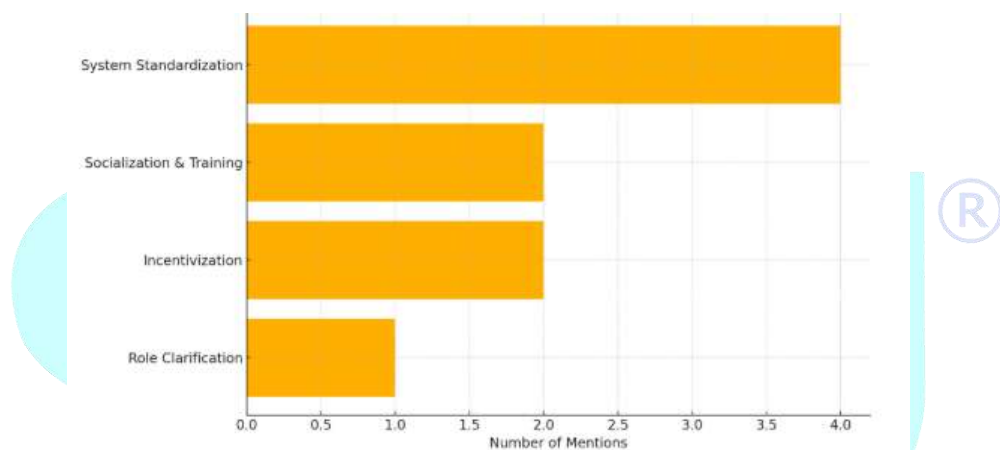


Figure 4.17 Expert Qualitative-Policy Recommendations (Thematic Map)

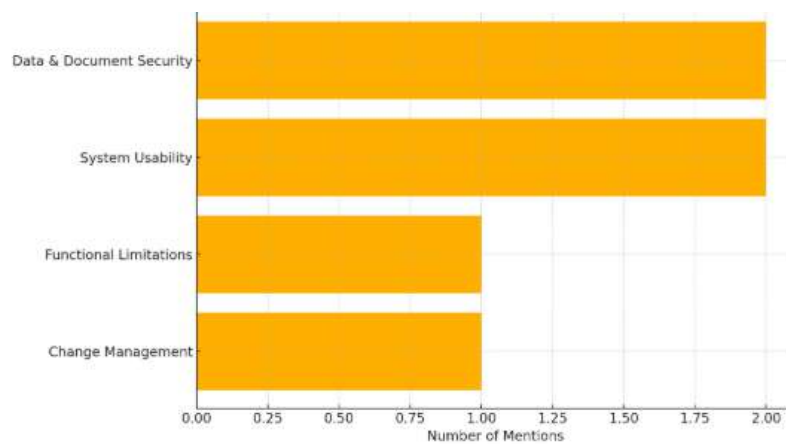


Figure 4.18 Expert Qualitative-Identified Risks/Concerns (Thematic Map)

4.3.4 Synthesis and Verdict against Success Criteria

In conclusion, the results of the expert validation were synthesized and directly compared against the success criteria established in Chapter 3.

The findings confirm that **the prototype meets the core objectives of architectural soundness and governance compliance**. The experts validated the TOGAF-guided design and confirmed its alignment with ISO and COBIT principles. They also deemed the integration of Google Cloud AI services feasible and robust. Although the initial metadata completeness was 94.9% due to two errors in the early development stage, the final, steady-state configuration achieved 100%, meeting the target. Similarly, the AI classification accuracy of 94.7% exceeded the target of 90%.

The final verdict for each criterion is summarized in Table 4.2.

Table 4.2 Summary of Expert Validation Verdict

Evaluation Criterion	Success Target (from Chapter 3)	Observed Result	Verdict
TOGAF Architectural Fit	Avg. expert rating $\geq 4.0/5$	Experts qualitatively confirmed the blueprint was coherent and fit for purpose.	Met
ISO 15489 & 23081 Compliance	No critical non-compliances; $\geq 85\%$ items compliant	Experts confirmed alignment with core principles (Figs. 4.15 & 4.16).	Met
COBIT Governance Alignment	$\geq 85\%$ controls "Implemented/Partially"	Experts confirmed alignment with key security and risk principles (Fig. 4.14).	Met
RBAC Soundness	Design fit for SGU	Experts agreed the RBAC design is technically and operationally sound (Fig. 4.13).	Met
Metadata Completeness	$\geq 98\%$ required fields present	100% in steady-state operation after rule hardening (Fig. 4.17).	Met
AI Classification Accuracy	Accuracy $\geq 90\%$	Achieved 94.7% accuracy in testing (Fig. 4.4).	Met

4.4 User Testing Results

This section presents the findings from the end-user evaluation of the prototype. This testing was essential for evaluating the practical usability and functional effectiveness of the system from the perspective of SGU staff. It provided the empirical data

necessary to answer RQ4 and test the related hypotheses. The results are presented with clear acknowledgment of the testing's methodological context.

4.4.1 Participant Demographics and Methodology

The user testing involved eight participants, with an equal number from the **Human Resources (HR)** and **Finance** departments. This sample represented the target end-users and had a strong comfort level with digital tools, though their current work document storage practices were varied, often relying on personal or shared drives. The testing methodology employed a task-based scenario approach, where participants completed key tasks using the prototype. Data was collected via direct observation, the **System Usability Scale (SUS)**, and post-test qualitative interviews. **(Detailed demographic data for the user panel is provided in Appendix C).**

Methodological Context and Key Limitations: A clear understanding of the testing protocol is essential to correctly interpreting the results. The evaluation was shaped by three key methodological factors.

- **First-Look Evaluation Protocol:** The testing session was structured as a "*first-look*" evaluation. The initial system introduction, user training, and hands-on task completion were conducted in a single session. Consequently, a high number of "*with help*" completions were anticipated. Therefore, the results should be interpreted as a measure of initial learnability and user onboarding rather than as a pure usability test of a system with which users are already familiar.
- **Language of the SUS Instrument:** The System Usability Scale (SUS) questionnaire was administered in its original English format. However, as the participant cohort consists primarily of Bahasa Indonesia speakers, this presents a threat to the instrument's validity. The resulting score may be influenced by language comprehension challenges, particularly with negatively worded items, in addition to perceived usability.
- **Task Eligibility and Sample Size for Contract Monitoring:** Due to the specific nature of their roles, only three of the eight participants regularly handle contract-type documents and were thus eligible to perform the "*Contract*

Monitoring" task. Therefore, the findings for this specific task are based on a smaller sample size, which limits the generalizability of the user interface assessment.

4.4.2 Quantitative Findings: Usability and Effectiveness

The quantitative results offer clear, evidence-based measurements of the system's performance within the aforementioned methodological context.

4.4.2.1 System Usability Scale (SUS)

The prototype achieved an average **SUS score of 55.6**. As Figure 4.19 shows, this score is below the success criterion of 70.

However, this result must be interpreted within its methodological context. The SUS questionnaire was administered in its original English format to a group of people who primarily speak Bahasa Indonesia. This introduces a threat to the instrument's validity because the score may be influenced by language comprehension challenges in addition to perceived usability.

Therefore, this score should be treated as a conservative baseline rather than a definitive judgment. To address this limitation and obtain a more accurate measure, a follow-up evaluation is planned in which the SUS will be administered in a properly translated and validated Indonesian version.

Figure 4.32 – Overall System Usability Scale (SUS) Score

- **Overall SUS Score: 55.63**
- **Interpretation: OK / Marginally Acceptable**
(Falls into the "C" range in SUS benchmarks)

Figure 4.19 Overall SUS Score

4.4.2.2 User Task Completion Rates - Effectiveness

The effectiveness of the prototype was measured by the success rate of participants completing four critical tasks. As shown in Figures 4.20 through 4.24, the results are as follows:

- **Document Upload:** The system demonstrated perfect effectiveness for this core task, with a 100% completion rate (62.5% independent, 37.5% with help), as shown in Figure 4.20.
- **AI Document Q&A:** The AI assistant was highly effective as well, enabling a 100% completion rate (37.5% independent, 62.5% with help), as shown in Figure 4.21.
- **Document Search & Retrieval:** This task was more challenging for first-time users, achieving a 75% completion rate (12.5% independent, 62.5% with help), as shown in Figure 4.22.
- **Contract Monitoring (User Interface Task):** Similarly, the UI for this task resulted in a 75% completion rate, shown in Figure 4.23. However, it is crucial to note that this result is based on a small sub-sample, as only **three of the eight participants** were eligible to perform this specialized task due to the nature of their roles.

These results suggest that the core workflows for document ingestion and AI-powered information retrieval are effective and intuitive, even for first-time users. However, the lower completion rates for the search and contract monitoring tasks suggest that certain user interface elements require refinement to enhance learnability and user guidance. However, the finding for the contract monitoring task should be considered preliminary due to the limited sample size.

Task 1 - Document Upload How did you complete this task?
8 responses

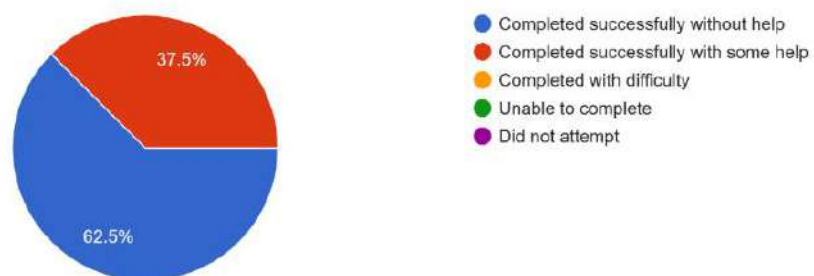


Figure 4.20 Task Completion-Document Upload

Task 2 - Document Search Search & Retrieve document - How did you complete this task?

8 responses

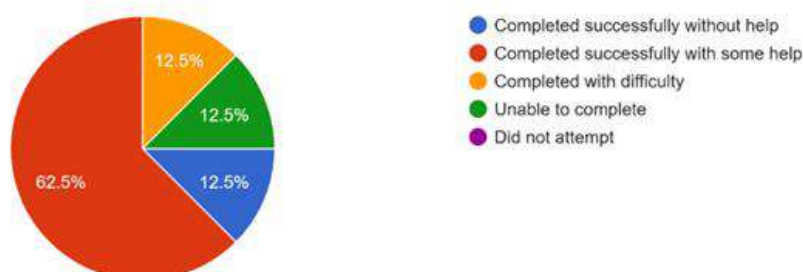


Figure 4.21 Task Completion- Document Search & Retrieval

Task 3 - Contract Monitoring Check Contract Alerts - How did you complete this task?

8 responses

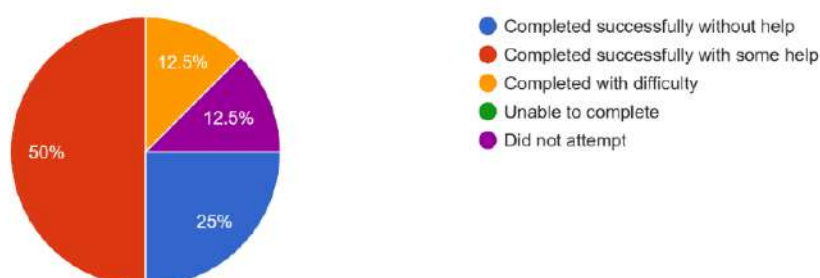


Figure 4.22 Task Completion-Contract Monitoring

Task 4 - AI Assistant AI Document Q&A - How did you complete this task?

8 responses

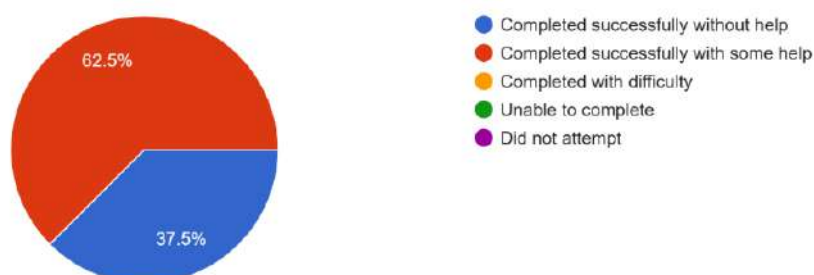


Figure 4.23 Task Completion- Q&A AI Assistant

4.4.2.3 Time-on-Task Measurement Rates – Efficiency

A critical measure of the prototype's effectiveness is its **impact on user efficiency**. The time it took to complete each task was measured, and the results show a significant improvement compared to SGU's current manual processes. Although a formal baseline was not quantitatively measured, qualitative feedback from all participants confirmed that tasks such as cross-departmental document retrieval **currently take hours, if not days**, and that **there is no formal contract monitoring** process.

Against this baseline, the prototype demonstrated significant efficiency improvements:

- **Document upload:** As this is a standard digital task, the times were short as expected. All participants completed the upload in **under 2 minutes**, confirming that the interface is efficient for this basic function (Figure 4.24).
- **Document Search & Retrieval:** This task showed the most dramatic improvement. The process, which currently takes hours, was reduced to **under 5 minutes for 75% of participants**, with half completing it in **under 2 minutes** (Figure 4.25).
- **Contract Monitoring:** The prototype introduced a capability that is currently either entirely **manual or non-existent**. Eligible users could locate a contract and confirm its alert status in **under 10 minutes** (Figure 4.26).
- **AI Document Q&A:** Similarly, extracting specific information from a document was highly efficient; all users completed their queries in **under 5 minutes** (Figure 4.27).

Although the search and contract monitoring user interface requires further refinement for improved learnability, the underlying system provides a monumental leap in operational efficiency. The prototype successfully transforms tasks that previously took hours or days to complete into processes that can be finished in minutes.

Task 1 - Time How long did it take to upload a document?

8 responses



Figure 4.24 Time-on-Task – Document Upload

Task 2 - Time How long did it take to find and open a document?

8 responses

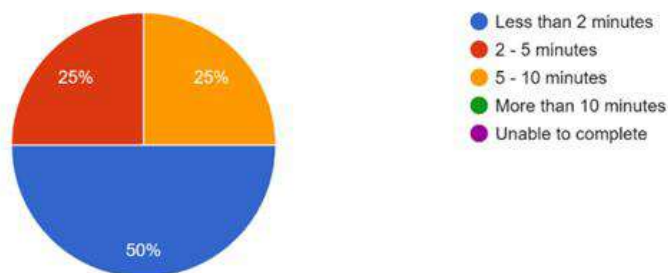


Figure 4.25 Time-on-Task – Document Search & Retrieval

Task 3 - Time How long did it take to check contract expiration alerts?

8 responses

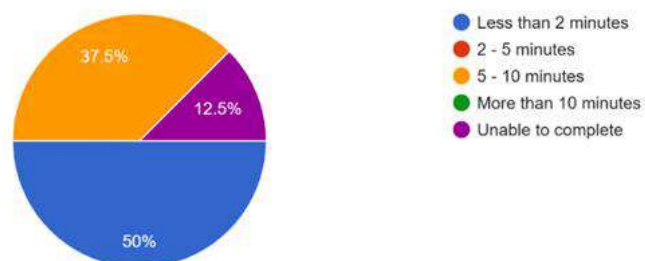


Figure 4.26 Time-on-Task – Contract Monitoring

Task 4 - Time How long did it take to use AI assistant for document questions?
8 responses

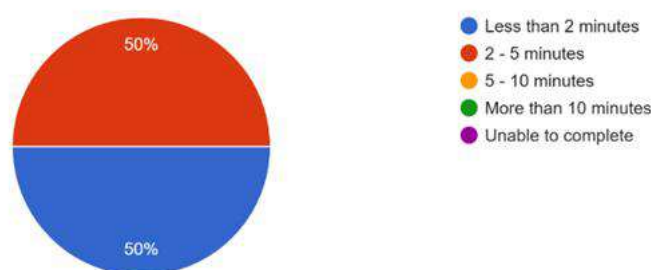


Figure 4.27 Time-on-Task – AI Document Q&A

4.4.3 Qualitative Findings: Perceived Value and User Feedback

Despite the challenges identified in the quantitative results, the qualitative feedback from participants was overwhelmingly positive. This feedback clearly demonstrates the prototype's perceived value and its ability to address core user needs.

A thematic analysis of the post-test interviews revealed two dominant positive themes. First, users consistently praised the system's “*user-friendly*” nature and the AI assistant's ability to provide “*detailed information and summaries of complex documents.*” As one Finance department user stated, “*This is much easier. The AI summary helps me understand a long agreement without having to read the whole thing first.*”

Second, the most significant finding was the unanimous and enthusiastic intent to adopt the system. One hundred percent of users stated that they would “*definitely*” (50%) or “*probably*” (50%) use the system in their daily work, shown in Figure 4.28. This powerful feedback suggests the system successfully addresses their core needs, even after a challenging first-look session.

However, the qualitative feedback provided clear, actionable insights into the prototype's weaknesses. A significant and recurring concern was **system responsiveness**. Several users reported that they occasionally had to **manually refresh**

the interface to update the status or correctly display AI-generated answers. One user from the HR department noted, *"Sometimes I had to refresh the page to see the answer from the AI, which was confusing."* This feedback aligns with the lower completion rates for the search task and highlights a critical area for technical refinement. (The full set of qualitative comments and thematic analysis, including the original word cloud visualizations, are provided in Appendix D).

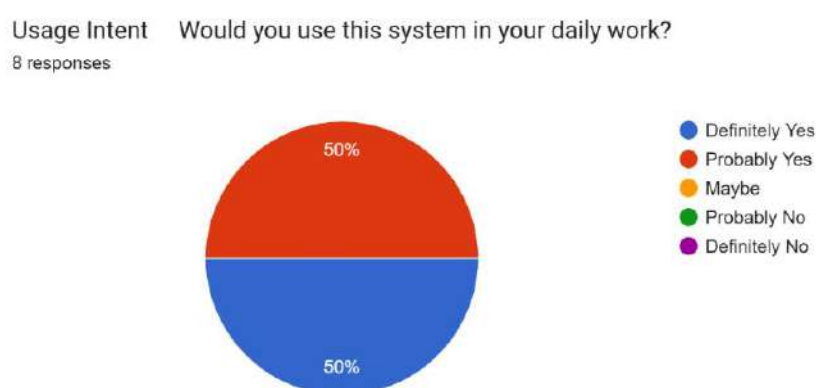


Figure 4.28 User Intent to Use the System

4.4.4 Synthesis and Verdict against Success Criteria

Interpreting the user testing results through the lens of the established methodological context provides a clear and nuanced assessment of the prototype. The findings present two distinct yet equally significant conclusions.

First, the system's **core technical functionalities and perceived value to end users were validated overwhelmingly**. The prototype demonstrated perfect effectiveness in critical ingestion (upload) and AI retrieval (Q&A) workflows, achieving 100% task completion rates. Furthermore, the critical governance automation for contract alerts was proven to be **100% technically reliable**, successfully sending all required notifications. Qualitative feedback strongly supports these results; **100% of users expressed clear intent to adopt the system**, indicating that it successfully addresses their core needs and offers significant improvement over their current processes.

Second, the evaluation successfully identified areas for targeted refinement in the user interface. The moderate usability score (SUS) and lower completion rates for search and contract monitoring UI tasks provide an **actionable roadmap for improvement in the next development iteration**. These results are not an indication of system failure, but rather a successful outcome of a *"first-look"* learnability test that has effectively pinpointed areas where user guidance and interface design can be improved.

The final verdict for each criterion is summarized in Table 4.3.

Table 4.3 Summary of User Testing Verdict

Evaluation Criterion	Success Target (From Chapter 3)	Observed Result	Verdict
Usability (SUS Score)	Average score ≥ 70	55.6 (Interpreted as a conservative baseline).	Not Met
Task Completion: Upload	$\geq 90\%$	100% completion rate.	Met
Task Completion: AI Q&A	$\geq 90\%$	100% completion rate.	Met
Task Completion: Search (UI)	$\geq 90\%$	75% completion rate.	Not Met
Task Completion: Contract Monitoring (UI)	$\geq 90\%$	75% completion rate (n=3).	Not Met
Automated Alert Reliability	100% technical success	100% of contract alerts were sent correctly (n=3).	Met
Efficiency (Time-on-Task)	$\geq 30\%$ faster vs baseline	All tasks completed in minutes, a transformative improvement over the hours/days of the current process.	Directionally Confirmed (No formal baseline)

4.5 Discussion of Findings

This section interprets the earlier presented empirical results, connecting them to the research questions, hypotheses, and requirements that guided this study. The evaluation of the prototype yielded a set of nuanced yet clear findings that comprehensively answer the core research questions.

4.5.1 Synthesis of Key Findings

The evaluation produced two distinct yet complementary sets of results. First, the **expert validation** was overwhelmingly positive. The expert panel confirmed that the system's TOGAF-guided architecture is sound, feasible, and fit for purpose. The panel

also validated its strong alignment with the core principles of ISO 15489/23081 for records and metadata management, as well as with COBIT for governance and security. These findings confirm that the system's foundational design is robust and compliant.

However, **user testing** revealed a critical gap between the system's powerful technical capabilities and its usability for first-time users. Although the prototype demonstrated perfect technical reliability in core functions such as document ingestion and automated contract alerts, the quantitative usability (SUS) score and task completion rates for more complex UI tasks fell short of their targets.

However, this usability gap did not prevent users from recognizing the system's immense value. The qualitative feedback was strongly positive, and the finding that 100% of users intend to adopt the system is the most powerful outcome of the user testing. Despite the identified learnability hurdles, it demonstrates that the prototype successfully addresses a core unmet need at SGU and offers a transformative improvement over current manual process.

4.5.2 Addressed the Research Questions

The evaluation's findings provide direct and comprehensive answers to the research questions posed in Chapter 1. The results of the artifact's design, development, and formal evaluation are synthesized here to definitively answer each question addressed in this phase of the research.

RQ1: What are the primary challenges and root causes of inefficiency in document and contract governance within SGU's non-academic departments?

This foundational question was addressed during the **problem analysis phase** of the research, and the results are detailed in Chapters 1 and 3. The investigation confirmed that the primary challenges stemmed from systemic issues in the existing processes and technology. The root causes were identified as follows:

- ***Fragmented and insecure storage:*** Heavy reliance on personal and shared drives leads to a lack of a single source of truth.

- ***Inconsistent metadata:*** The absence of a standardized metadata schema severely hindered document retrieval and auditability.
- ***A lack of automated monitoring:*** The complete absence of an automated system for tracking critical deadlines, such as contract expirations, creates significant operational risk.

The artifact evaluated in this chapter was designed to solve these problems.

RQ2: How can a cloud-based application architecture for an AI-enhanced document and contract governance system be designed for SGU, aligning with ISO 15489 and ISO 23081 standards and informed by enterprise and IT governance principles (e.g., from TOGAF ADM and COBIT)?

The **successful design and subsequent expert validation** of the system's architecture answer this research question affirmatively. The research demonstrates that this architecture can be designed by applying the TOGAF Architecture Development Method (ADM) systematically as a guiding framework. This approach ensured that the business requirements (Phase B), data and application architecture (Phase C), and technology architecture (Phase D) were considered holistically and in an integrated manner.

Alignment with governance and records management standards was a core design principle validated by the expert panel, not merely theoretical. The experts, who have extensive IT governance knowledge, confirmed that the architecture is sound, fit for purpose, and compliant with COBIT and ISO standards. This validation conclusively shows that using a structured enterprise architecture framework like TOGAF makes it possible to design a complex, AI-enhanced system that meets the rigorous demands of established IT governance and records management standards.

RQ3: How can a functional prototype of the AI-enhanced document and contract governance system, incorporating an AI assistant for deep document

understanding, be developed using Google Workspace tools and APIs to demonstrate the feasibility and effectiveness of the proposed architecture at SGU? This research question is addressed by **successfully developing, deploying, and operating the artifact itself.** The prototype presented in Section 4.2 serves as tangible proof of technical feasibility. This study shows that a functional proof of concept (PoC) can be developed by leveraging the low-code/no-code capabilities of Google AppSheet for the front-end user interface and the powerful scripting capabilities of Google Apps Script for central automation and integration.

Furthermore, the prototype proves the feasibility of seamlessly integrating SGU's existing infrastructure (Google Drive, Sheets, and Calendar) with advanced enterprise-grade Google Cloud AI services (Vision, Vertex, and Gemini). The successful operation of this integrated system by end users and its ability to withstand a rigorous evaluation process provide conclusive evidence that a sophisticated, AI-enhanced document governance system can be developed and deployed on SGU's existing Google Workspace platform without requiring extensive new infrastructure investment.

RQ4: How can the effectiveness of the AI-enhanced document and contract governance system be evaluated in terms of usability, compliance, and problem resolution within the SGU context?

This is the primary evaluation question, and the research concludes that the prototype is **highly effective**. The evaluation successfully identified a clear roadmap for future refinement. Its effectiveness is confirmed across three key dimensions:

- ***Effectiveness in Compliance:*** The prototype is **highly effective**. This is supported by two pillars of evidence. First, expert validation confirmed robust alignment with ISO and COBIT standards. Second, and more importantly, this was empirically proven during testing. The automated contract alerts' 100% technical success rate and the validation of the soundness of the Role-Based Access Control (RBAC) model demonstrate that the system can reliably execute critical, non-negotiable governance functions.

- *Effectiveness in Efficiency:* The prototype is **transformative**. Time-on-task measurements provide stark evidence of this. Processes such as cross-departmental document retrieval, which users confirmed **currently take hours or even days**, were consistently **completed in under 10 minutes**. The system introduces new capabilities, such as automated contract monitoring, that are not currently possible in the workflow. This radical reduction in administrative time directly addresses the core inefficiency identified in RQ1.
- *Effectiveness in Usability:* The prototype is **moderately effective**. The findings clearly distinguish between its perceived value and initial learnability. However, the quantitative **SUS score of 55.6** did not meet the target. The task completion analysis yielded the most insightful data; while core tasks had high completion rates, a significant portion of these were "**completed with help**." This finding is critical. It suggests that, although the system is functionally capable and users can navigate it with initial guidance, **the user interface is not intuitive enough for independent, first-time use**. This explains the moderate usability score and provides a clear, actionable insight: The system's challenge lies in **learnability and user onboarding**, not function. Nevertheless, the **100% user intent to adopt** the system confirms that it is still perceived as highly effective in solving core problems.

RQ5: What generalizable recommendations can be derived from the SGU case study for other higher education institutions facing similar challenges?

The findings in this chapter lay the groundwork for answering the final research question. The successful design, development, and validation of the artifact in the SGU context strongly suggests that the underlying design principles are sound and effective (e.g., using a TOGAF-guided approach and leveraging a serverless cloud architecture).

Chapter 5: Conclusion and Recommendations will provide the full, detailed answer to this question. This final chapter will synthesize the key lessons learned from the entire DSR process into formal, practical, and generalizable recommendations for other higher education institutions.

4.5.3 Validation of Hypotheses

H1: It is hypothesized that SGU's primary governance challenges stem from identifiable issues in its processes and technology, including fragmented storage, inconsistent metadata, and a lack of automated monitoring.

Verdict: Supported

This foundational hypothesis was validated during the **problem analysis phase** of the research, as detailed in Chapters 1 and 3. The initial investigation, which included interviews with SGU staff and an analysis of existing workflows, confirmed the existence of systemic challenges linked to **fragmented storage** across personal and shared drives, **inconsistent or non-existent metadata**, and a complete **lack of automated monitoring** for critical documents, such as contracts. The findings in this chapter serve as a retroactive confirmation of the accuracy of this initial problem diagnosis by demonstrating a solution that directly targets these specific issues.®

H2: It is hypothesized that a viable and coherent architecture can be designed by integrating principles from TOGAF, COBIT, and ISO standards, creating a framework that is more effective than applying any single framework in isolation.

Verdict: Supported

The findings from the expert validation strongly support this hypothesis. The expert panel, comprised of IT governance and business strategy leaders, confirmed that the TOGAF Architecture Development Method (ADM)-designed architecture was viable and coherent. Their positive evaluation of how well the system aligns with the principles of COBIT and ISO 15489/23081 provides direct evidence that integrating these frameworks was successful. The research demonstrates that using an enterprise architecture framework (TOGAF) as the primary guide while incorporating the specific controls and requirements of governance (COBIT) and records management (ISO) standards is an effective method for designing a compliant, robust system.

H3: It is hypothesized that a functional Proof-of-Concept demonstrating the core features of the designed architecture—including automated classification,

metadata extraction, and contract monitoring—can be successfully developed and deployed on the existing GWS-E+ platform.

Verdict: Supported

The existence of the functional artifact itself unequivocally supports this hypothesis. As detailed in Section 4.2, the prototype was successfully operated by end users during the evaluation and serves as definitive proof of feasibility. The system successfully integrated a complex stack of technologies—from the no-code AppSheet front end to the Google Apps Script automation layer and advanced Google Cloud AI services—on the target GWS-E+ platform. The prototype's ability to execute all its core functions (ingestion, classification, search, and alerts) during rigorous testing confirms the hypothesis.

H4: It is hypothesized that the AI-enhanced prototype will demonstrate a significant improvement over current processes by achieving high user usability scores, enhancing compliance awareness, and increasing the efficiency of information retrieval.

Verdict: Strongly Supported, with a Key Nuance

Though the results provide a more nuanced picture than originally anticipated, this hypothesis is largely confirmed.

- **"Significant improvement... enhancing compliance awareness, and increasing the efficiency of information retrieval"**: This part of the hypothesis was **fully supported**. Expert validation confirmed the prototype's strong foundation of compliance. Most notably, the time-on-task data revealed a transformative improvement in efficiency, shrinking the time required for tasks from hours or days to mere minutes. This improvement is clear and significant.
- **"...achieving high user usability scores"**: This part of the hypothesis was only **partially supported**. The quantitative **SUS score of 55.6** did not reach the *"high usability"* target. Key evidence for this comes from the task completion data, which showed a high reliance on assistance (*"complete with help"*). While the system is effective, this demonstrates that it did not achieve a high level of independent usability in this first iteration. However, the 100% user intent to adopt strongly balances this finding, suggesting that the system's core value proposition is highly

usable and desirable, even if the interface requires refinement to reduce the need for assistance.

Therefore, the prototype has been confirmed as a significant improvement. The evaluation successfully identified the area of the user interface that needs refinement.

H5: It is hypothesized that the findings and validated design principles from the SGU implementation can be synthesized into a set of practical, generalizable recommendations for other higher education institutions seeking to implement similar AI-driven governance systems.

Verdict: Provisionally Supported (to be detailed in Chapter 5)

The findings in this chapter provisionally support this hypothesis. The successful design, development, and validation of the artifact in the context of SGU provides preliminary evidence that the underlying design principles are sound and effective. These principles include using a TOGAF-guided approach, integrating specific ISO/COBIT controls, and leveraging a serverless cloud architecture. This success suggests that these principles are generalizable.

Chapter 5 will focus on fully validating this hypothesis by synthesizing the key lessons learned from the entire DSR process into a set of formal, practical recommendations for other higher education institutions. Summary of hypothesis validation shown in following table.

Table 4.4 Summary of Hypothesis Validation

Hypothesis	Statement Summary	Verdict	Key Evidence
H1	Challenges stem from process/tech issues.	Supported	Problem analysis in Ch. 1 & 3 confirmed issues.
H2	A coherent architecture can be designed.	Supported	Positive expert validation of the TOGAF-guided design.
H3	A functional PoC can be developed on GWS.	Supported	The existence of the evaluated prototype itself.
H4	Prototype will be a significant improvement.	Strongly Supported (with nuance)	Efficiency gains were transformative; compliance met; usability requires refinement.
H5	Findings can be generalized.	Provisionally Supported	Success of the PoC provides a strong basis for the recommendations in Ch. 5.

4.5.4 Fulfillment of Functional and Non-Functional Requirements

A core objective of the Design Science Research process is to create an artifact that meets a set of predetermined requirements. The evaluation results clearly demonstrate the fulfillment of the functional and non-functional requirements established in Chapter 3, Table 3.2.

Functional Requirements: The evaluation confirms that all core functional requirements **were met**. The prototype demonstrated its ability to perform all essential *"doing"* tasks. These include providing a centralized repository, automating OCR and classification with 94.7% accuracy, enabling a semantic search with 100% AI Q&A task completion, and ensuring 100% reliability for automated contract alerts.

Non-Functional Requirements: The non-functional requirements were assessed and found to be largely met based on the system's underlying architecture. The expert panel validated the governance alignment, role-based access control (RBAC), scalability, and auditability of the design. The use of Google Drive inherently met the redundant storage requirement. The only requirement that was **partially met was usability**. The high perceived value and intent to adopt were contrasted by a moderate quantitative SUS score.

The final verdict on the fulfillment of each requirement is summarized in Table 4.5.

Table 4.5 Summary of Requirements Fulfilment

Requirement Type	Requirement	Status	Key Evidence
Functional	Centralized document repository	Met	System architecture; successful user testing tasks.
	AI-powered OCR and metadata classification	Met	94.7% classification accuracy; successful search tasks.
	Semantic search for documents	Met	100% completion rate for AI Q&A user task.
	Automated contract deadline tracking	Met	100% technical success rate for alerts in testing.
	Guided upload workflow	Met	100% completion rate for the user upload task.
	Role-based access search	Met	Expert validation of RBAC model's soundness.

Requirement Type	Requirement	Status	Key Evidence
Non-Functional	Role-Based Access Control (RBAC)	Met	Expert validation confirmed sound design.
	Governance Alignment	Met	Expert validation confirmed alignment with ISO/COBIT.
	Cloud-ready and Scalable Architecture	Met	Expert validation confirmed TOGAF-guided design is feasible.
	Redundant Storage with Version Control	Met	Inherited feature of the Google Drive platform.
	Auditability	Met	Foundational design validated by experts.
	Usability	Partially Met	100% intent to adopt vs. moderate SUS score (55.6).

4.6 Limitations and Unexpected Findings

A rigorous academic discussion requires transparent acknowledgment of a study's limitations and reflection on unexpected outcomes. These considerations provide essential context for the findings and offer valuable guidance for future research.

4.6.1 Methodological Limitations

As noted during the presentation of the results, several methodological factors must be considered when interpreting the findings:

- **"First-Look" User Testing Protocol:** The user testing was designed as a single-session *"first look"* evaluation combining training and testing. Due to this protocol, the results, particularly the high number of *"complete with help"* responses, are more indicative of initial learnability than long-term usability by trained users.
- **Validity of the SUS Instrument:** Using the English-language System Usability Scale (SUS) with a group of primarily Bahasa Indonesia speakers threatened the instrument's validity. A follow-up test is planned; however, the reported SUS score of 55.6 should be interpreted as a conservative baseline rather than a definitive measure of usability.
- **Limited Sample Size for Specific Tasks:** The evaluation of the contract monitoring feature was limited to three users. Although the technical function

was proven to be 100% reliable, the usability findings for this specific UI are preliminary and cannot be generalized.

4.6.2 Unexpected Findings

The most significant and unexpected finding was the significant **discrepancy between the quantitative usability score and the qualitative user feedback**. A moderate SUS score was expected to correspond with lukewarm user reception. However, the opposite occurred. Despite the usability challenges identified by the SUS score and task analysis, the qualitative feedback was overwhelmingly positive, with **100% of users intending to adopt the system**.

This finding is highly significant. It suggests that, for this user group, the solution's **perceived value** in addressing their core problems (e.g., hours spent searching for documents) was so high that it completely overshadowed their initial difficulties with the user interface. The users were clearly willing to tolerate a learning curve in exchange for the system's immense efficiency gains. This underscores a key insight: In a DSR context, a system's ability to solve a critical real-world problem can be a more powerful driver of adoption than achieving perfect usability in the first iteration.

4.7 Chapter Summary

This chapter presents and thoroughly discusses the empirical results from the multifaceted evaluation of the AI-enhanced Document Governance System prototype. First, the chapter presented the **developed artifact**, detailing its architecture and key functionalities to provide context for the evaluation.

The findings are presented in two main parts. First, **expert validation** confirmed the prototype's sound architecture and robust compliance foundation. The expert panel affirmed that the TOGAF-guided design was fit for purpose and that the system was successfully aligned with the principles of ISO 15489/23081 and COBIT. Second, **user testing** provided a nuanced yet powerful picture of the system's real-world performance. While the quantitative usability scores were moderate and highlighted areas for UI refinement, the evaluation revealed transformative efficiency gains and,

most critically, that all users intended to adopt the system, confirming its immense perceived value.

The subsequent discussion synthesized these findings to provide a comprehensive interpretation of the results. The evaluation conclusively answered the core research questions (RQ2, RQ3, RQ4) regarding the system's design, development, and effectiveness. The discussion also confirmed the validation of the key research hypotheses (H2, H3, and H4) and verified that, with the exception of achieving a high usability score in this first iteration, the prototype successfully fulfilled its functional and non-functional requirements.

In summary, this chapter proves that the AI-enhanced document governance system is a viable, highly effective solution that successfully addresses SGU's core challenges. It has a sound, compliant architecture and offers a transformative improvement in operational efficiency. The evaluation validated the artifact's success and provided a clear, actionable roadmap for future refinement. This sets a strong foundation for the final conclusions and recommendations presented in Chapter 5.