AI-Enhanced Email Knowledge Recovery Benchmark

Executive Summary

This whitepaper presents a benchmark analysis of three Retrieval Augmented Generation (RAG) solutions for email knowledge recovery: mxHERO Mail2Cloud Advanced (mxMCP), Box AI Studio, and a generic Vector Search implementation. Testing was conducted using a dataset of approximately 1,730 emails, with all solutions connected via Model Context Protocol (MCP) to Claude 4 Sonnet.

Key Finding: mxHERO Mail2Cloud Advanced achieved 100% accuracy across all test categories, outperforming Box AI Studio (46.2%) and Vector Search (30.8%) overall accuracy. This result is primarily due to data and search optimizations tailored for email data that afforded Mail2Cloud Advanced higher quality retrieval for AI systems specific to that data type.

Methodology

Test Environment

- **Dataset**: ~1,730 emails collected via mxHERO Mail2Cloud
- **LLM Model**: Claude 4 Sonnet
- **Connection Method**: Model Context Protocol (MCP) clients
- **Test Categories**: 5 distinct categories with 13 total queries

Solutions Tested

1. mxHERO Mail2Cloud Advanced (mxMCP)

A purpose-built email knowledge recovery solution with specialized optimizations:

- Email content deduplication to reduce redundancy from replies/forwards

- Metadata-driven search scope reduction using From, To, Date, Subject fields

- AI-powered query parsing for structured metadata searches

- Statistical and chronological search adaptations

- Original email preservation in Box with secure Box share links for source verification

URL: https://www.mxhero.com/advanced-ai

2. Box AI Studio

A versatile RAG solution for all content types, easily deployed and configured:

- Emails converted to PDFs with metadata attached

- Custom prompt engineering to identify email content

- Configuration: Temperature = 0, Claude 4 Sonnet

URL: https://www.box.com/ai/ai-studio

3. Vector Search

A generic vector search implementation:

- Email content and headers stored together without metadata extraction

- Cosine similarity distance metric

- OpenAI text-embedding-3-small model (1,536 dimensions)

Test Design

This study measured RAG performance with 13 queries across five categories of questions, primarily for illustrative purposes. Additional questions within each of the category types exhibited results consistent with those presented here. In other words, mxMCP consistently outperformed the other methods for 'statistical' queries, whereas queries with specific terms, unique to the target emails (needle-in-the-haystack), performed well for all methods.

Deficiencies

The data set comprises mailing list emails embedded with mock emails simulating different use cases across diverse verticals. The test corpus is not a faithful representation of real emails. Furthermore, most emails in the dataset do not have reply or forwarded content in their message bodies. This dramatically reduces the amount of redundant content that might be retrieved and consequently dilute the solution's performance. Mail2Cloud Advanced mitigates embedded redundant content. In a real-world scenario, mxMCP would perform even better compared to the other solutions measured in this report because of its data optimizations.

Results Analysis

Overall Performance

Solution	Correct Answers	Total Questions	Accuracy
mxMCP	13	13	100%
Box AI	6	13	46.2%
Vector Search	4	13	30.8%

Graph of Results

Correct responses out of 13 questions
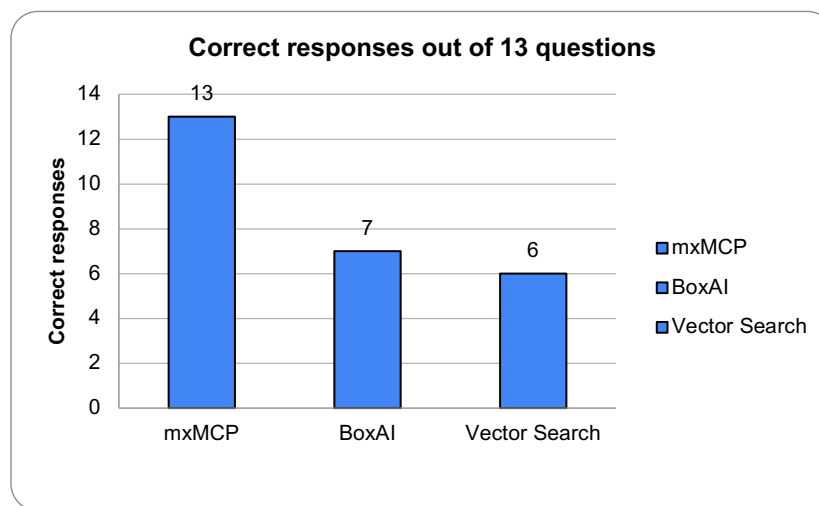
Category-by-Category Analysis

1. Statistical Queries (3 questions)

Questions focused on quantities and date ranges.

Solution	Performance	Analysis
mxMCP	3/3 (100%)	Custom recognition of statistical queries enabled specialized database counts
Box AI	0/3 (0%)	Unable to perform statistical operations on email data
Vector Search	1/3 (33%)	Limited success with basic counting

Example Query: "How many emails are from Bigcorp?"

2. Date Targeting (2 questions)

Queries requiring specific date range filtering.

Solution	Performance	Analysis
mxMCP	2/2 (100%)	Metadata extraction enabled precise date-based filtering
Box AI	1/2 (50%)	Partial success with date-specific content
Vector Search	0/2 (0%)	Unable to effectively filter by date ranges

Example Query: "List the subjects of all emails referencing OpenAI from the first week of July 2024."

3. Attachments (2 questions)

Questions about content within email attachments.

Solution	Performance	Analysis
mxMCP	2/2 (100%)	Full attachment indexing and search
Box AI	2/2 (100%)	Strong performance with PDF-converted attachments
Vector Search	2/2 (100%)	Semantic matching successful for specific terms

All solutions performed well due to specific keywords ("Project Kronos", "proposal to Bigcorp") present in the limited email set.

4. General Discovery (5 questions)

Broad email discovery queries.

Solution	Performance	Analysis
mxMCP	5/5 (100%)	Leveraged chronological and statistical optimizations

Solution	Performance	Analysis
Box AI	3/5 (60%)	Moderate success with general queries
Vector Search	2/5 (40%)	Limited to basic semantic matching

Notable Success: mxMCP's ability to handle "What is the latest email from Michael Spencer?" through chronological search adaptation.

5. Deep Analysis (1 complex query)

Complex analytical query requiring comprehensive email analysis.

Solution	Performance	Key Metric
mxMCP	1/1 (100%)	Found all 15 relevant emails
Box AI	0/1 (0%)	Found only 5 relevant emails
Vector Search	0/1 (0%)	Found only 6 relevant emails

Query: "Access the email record to and from our client Bigcorp and build a summary data sheet showing customer feedback…"

This category highlighted the critical importance of comprehensive retrieval, as LLMs can generate convincing but incomplete results when provided with partial data.

Key Differentiators

mxHERO Mail2Cloud Advanced Advantages

1. **Email-Specific Optimizations**
   - Content deduplication reduces noise from email threads
   - Metadata preservation enables precise filtering
   - Statistical and chronological search modes
2. **Scalability**
   - Metadata filtering ensures effectiveness regardless of repository size
   - Optimized for LLM context window limits
3. **Source Verification**
   - Preserved original emails with Box share links
   - Protection against user deletion

Box AI Studio Strengths

- Rapid deployment (minutes)
- Versatile for multiple content types
- Strong attachment handling through PDF conversion

Vector Search Limitations

- No metadata extraction
- Limited date/statistical capabilities
- Basic similarity matching only

AI Enhancement via MCP

The Model Context Protocol represents a significant advancement in AI-email integration:

- **Standardized Integration**: Any AI solution can leverage specialized data acquisition tools
- **Intelligent Query Strategies**: Claude 4 Sonnet demonstrated sophisticated multi-query approaches, automatically refining searches based on initial results
- **Adaptive Search**: When queries yielded insufficient data, the model autonomously broadened search parameters

Source Data and Study Replication

All data used in this study is publicly available. Furthermore, demo accounts for mxMCP can be freely acquired to test Mail2Cloud advanced preloaded with the dataset. See the following links:

Data used in this study is publicly available at available on GitHub at:
https://github.com/mxaiorg/eml-demo-1

- Result Excel - https://github.com/mxaiorg/eml-demo-1/tree/main/mcp-1

Demo accounts for Mail2Cloud Advanced are available at:
https://mxhero.helpjuice.com/en_US/mxhero-ai/demo-account-for-ai-testing

Appendix A: Questions & Answers

Full results at: https://github.com/mxaiorg/eml-demo-1/tree/main/mcp-1

	Benchmark Tests			
	Model: Claude 4 Sonnet			
		Correct Answer		
	Query	**mxMCP**	**BoxAI**	**Vector Search**
	Statistical			
1	How many emails are there?	Yes	No	No
2	How many emails are from Bigcorp?	Yes	No	No
3	How many emails are there from Sarah Jenkins, the lawyer, from June 27, 2023 and June 29 2023?	Yes	No	Yes
	Date Targeting			
4	What is the lead story in the email from Michael Spencer on February 26, 2024?	Yes	Yes	No
5	List the subjects of all emails referencing OpenAI from the first week of July 2024.	Yes	No	No
	Attachments			
6	Summarize the final analysis document regarding Project Kronos.	Yes	Yes	Yes
7	When did we send a proposal to Bigcorp?	Yes	Yes	Yes
	General Discovery			
8	Who initiated the conversation regarding Project Titan?	Yes	Yes	Yes
9	What aspects of our service was Bigcorp interested in?	Yes	Yes	Yes
10	What is the latest email from Michael Spencer?	Yes	No	No
11	What project was Sarah and Maria working on around April of 2025?	Yes	Yes	Yes
12	How many support tickets where there from us to Bigcorp?	Yes	Yes	No
	Deep Analysis			

13	Access the email record to and from our client Bigcorp and build for me a summary data sheet showing customer feedback. In the summary (on one page), break down by accolades/compliments and a second section that highlights any feedback, support tickets, or insights that might allude to customer satisfaction risks.	Yes	No	No
	Total Correct	13	7	6