Multi-Source RAG for Technical Support

TaskFlow Lite (Fictional Product)

Srutileka Suresh

February 2026

1 Problem Statement

Build a Retrieval-Augmented QA system that answers customer support questions about a fictional software product (TaskFlow Lite) by retrieving evidence from **three distinct sources**: product documentation, customer forums, and technical blog posts. The system must implement source-specific chunking, multi-source retrieval with weighting, reranking, contradiction handling, and logging.

2 Product Overview: TaskFlow Lite

TaskFlow Lite is a lightweight project/task management tool. Users can manage projects (with a task limit), attach files to tasks (with file type + size limits), and use an API (with key expiry and rate limits). The knowledge sources intentionally include cases where community forums or blogs conflict with official docs.

3 Dataset (Custom made)

- **Documentation (.md):** Authoritative product rules such as attachment limits, export formats, API key expiry, and error codes.
- **Customer Forums(.jsonl):** User questions and replies; may be noisy or contradictory.
- **Technical Blogs(.md):** How-to posts and best practices; sometimes outdated.

4 Chunking Strategy

Different sources have different structure, so chunking is tuned per source:

- **Docs:** Larger, structured chunks to preserve bullet lists (limits, error codes) as a single coherent unit.
- **Blogs:** Medium chunks to keep narrative explanations together.
- **Forums:** Small chunks per question or per reply to keep distinct opinions separated (and to support contradiction detection).

5 Retrieval (Multi-Source + Weighting)

5.1 Why separate retrieval per source?

Three separate FAISS indexes are built (`docs`, `blogs`, `forums`). At query time, the system retrieves top-k from each index to ensure every source can contribute evidence.

5.2 Source weighting

After retrieving from each source, the system applies source weights (docs prioritized):

$$\texttt{weighted\_score} = \texttt{similarity\_score} \times w_{source}$$

Weights used in the run: `docs=1.2`, `blogs=1.0`, `forums=0.8` :contentReference[oaicite:2]index=2.

6 Reranking (Cross-Encoder)

The system uses a two-stage ranking pipeline:

1. **Bi-encoder retrieval** (fast): retrieve a shortlist of candidates via embedding similarity (FAISS).
2. **Cross-encoder reranking** (precise): rerank only the top candidates

This design balances speed (vector search) and relevance (token-level interaction in cross-encoder).

7 Contradiction Handling

Contradictions can occur between sources (e.g., size limits, expiry durations, rate limits). The system:

- extracts numeric claims (e.g., `25MB`, `60 days`, `300/hour`) from top evidence,
- flags conflicts when multiple values appear for the same category,
- resolves by a reliability policy: **Docs** > **Blogs** > **Forums**.

8 Logging

Each query is logged to a JSONL file including:

- query string,
- top candidates pre- and post-rerank,
- sources used and scores,
- detected contradictions and the selected reliability policy.

9 Performance Analysis

We compare retrieval quality before reranking (`pre`) vs after reranking (`post`) using file-level "gold" targets for the synthetic dataset :contentReference[oaicite:4]index=4.

9.1 Summary Metrics

From `metrics.json` :contentReference[oaicite:5]index=5:

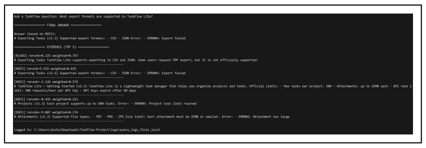
Metric	Pre-rerank	Post-rerank
Hit@5	0.80	1.00
Hit@1	0.70	0.40

Table 1: Retrieval performance before vs after reranking (10 queries).

Interpretation:

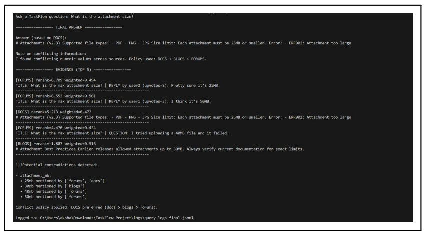
- Hit@5 improves from 0.80 to 1.00, meaning reranking + multi-source merging increased the chance that the correct file appears in the top-5.
- Hit@1 decreases from 0.70 to 0.40 in this run. This happens because the cross-encoder sometimes ranked short, direct forum/blog snippets above the authoritative doc chunk (even though the doc chunk remained in the top results). The system mitigates this via a **source reliability policy** during answer selection.

10 Query Result Screenshots



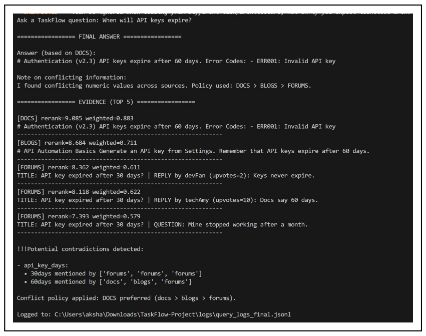
```
Ask a TaskFlow question: What export formats are supported in TaskFlow Lite?

================ FINAL ANSWER ================

Answer (based on DOCS):
# Exporting Tasks (v2.3) Supported export formats: - CSV - JSON Error: - ERR004: Export failed

================ EVIDENCE (TOP 5) ================

[BLOGS] rerank=8.221 weighted=0.767
# Exporting Tasks TaskFlow Lite supports exporting to CSV and JSON. Some users request PDF export, but it is not officially supported.
----------------------------------------------------------
[DOCS] rerank=5.539 weighted=0.635
# Exporting Tasks (v2.3) Supported export formats: - CSV - JSON Error: - ERR004: Export failed
----------------------------------------------------------
[DOCS] rerank=-1.126 weighted=0.575
# TaskFlow Lite — Getting Started (v2.3) TaskFlow Lite is a lightweight task manager that helps you organize projects and tasks. Official Limits: - Max tasks per project: 500 - Attachments: up to 25MB each - API rate l
imit: 300 requests/hour per API key - API keys expire after 60 days
----------------------------------------------------------
[DOCS] rerank=-8.435 weighted=0.221
# Projects (v2.3) Each project supports up to 500 tasks. Error: - ERR005: Project task limit reached
----------------------------------------------------------
[DOCS] rerank=-9.087 weighted=0.276
# Attachments (v2.3) Supported file types: - PDF - PNG - JPG Size limit: Each attachment must be 25MB or smaller. Error: - ERR002: Attachment too large
----------------------------------------------------------

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(a) Q1 Output



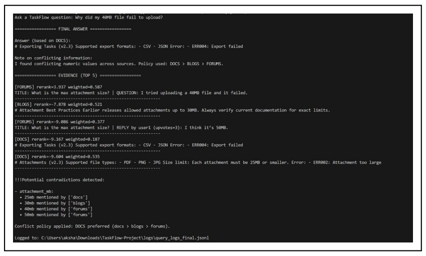
```
Ask a TaskFlow question: What is the attachment size?

================ FINAL ANSWER ================

Answer (based on DOCS):
# Attachments (v2.3) Supported file types: - PDF - PNG - JPG Size limit: Each attachment must be 25MB or smaller. Error: - ERR002: Attachment too large

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[FORUMS] rerank=6.709 weighted=0.494
TITLE: What is the max attachment size? | REPLY by user2 (upvotes=8): Pretty sure it's 25MB.
----------------------------------------------------------
[FORUMS] rerank=6.553 weighted=0.501
TITLE: What is the max attachment size? | REPLY by user1 (upvotes=3): I think it's 50MB.
----------------------------------------------------------
[DOCS] rerank=5.213 weighted=0.472
# Attachments (v2.3) Supported file types: - PDF - PNG - JPG Size limit: Each attachment must be 25MB or smaller. Error: - ERR002: Attachment too large
----------------------------------------------------------
[FORUMS] rerank=4.470 weighted=0.434
TITLE: What is the max attachment size? | QUESTION: I tried uploading a 40MB file and it failed.
----------------------------------------------------------
[BLOGS] rerank=-1.807 weighted=0.516
# Attachment Best Practices Earlier releases allowed attachments up to 30MB. Always verify current documentation for exact limits.
----------------------------------------------------------

!!!Potential contradictions detected:

- attachment_mb:
  • 25mb mentioned by ['forums', 'docs']
  • 30mb mentioned by ['blogs']
  • 40mb mentioned by ['forums']
  • 50mb mentioned by ['forums']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(b) Q2 Output



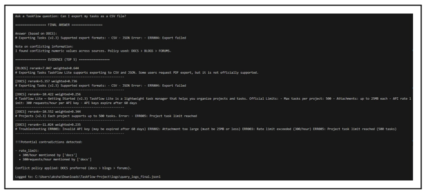
```
Ask a TaskFlow question: When will API keys expire?

================= FINAL ANSWER =================

Answer (based on DOCS):
# Authentication (v2.3) API keys expire after 60 days. Error Codes: - ERR001: Invalid API key

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================= EVIDENCE (TOP 5) =================

[DOCS] rerank=9.085 weighted=0.883
# Authentication (v2.3) API keys expire after 60 days. Error Codes: - ERR001: Invalid API key
----------------------------------------------------------
[BLOGS] rerank=8.684 weighted=0.711
# API Automation Basics Generate an API key from Settings. Remember that API keys expire after 60 days.
----------------------------------------------------------
[FORUMS] rerank=8.362 weighted=0.611
TITLE: API key expired after 30 days? | REPLY by devFan (upvotes=2): Keys never expire.
----------------------------------------------------------
[FORUMS] rerank=8.118 weighted=0.622
TITLE: API key expired after 30 days? | REPLY by techAmy (upvotes=10): Docs say 60 days.
----------------------------------------------------------
[FORUMS] rerank=7.393 weighted=0.579
TITLE: API key expired after 30 days? | QUESTION: Mine stopped working after a month.
----------------------------------------------------------

!!!Potential contradictions detected:

- api_key_days:
  • 30days mentioned by ['forums', 'forums', 'forums']
  • 60days mentioned by ['docs', 'blogs', 'forums']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(c) Q3 Output



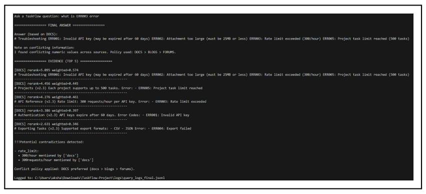
```
Ask a TaskFlow question: Why did my 40MB file fail to upload?

================ FINAL ANSWER ================

Answer (based on DOCS):
# Exporting Tasks (v2.3) Supported export formats: - CSV - JSON Error: - ERR004: Export failed

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[FORUMS] rerank=3.937 weighted=0.587
TITLE: What is the max attachment size? | QUESTION: I tried uploading a 40MB file and it failed.
----------------------------------------------------------
[BLOGS] rerank=-7.878 weighted=0.521
# Attachment Best Practices Earlier releases allowed attachments up to 30MB. Always verify current documentation for exact limits.
----------------------------------------------------------
[FORUMS] rerank=-9.086 weighted=0.377
TITLE: What is the max attachment size? | REPLY by user1 (upvotes=3): I think it's 50MB.
----------------------------------------------------------
[DOCS] rerank=-9.167 weighted=0.187
# Exporting Tasks (v2.3) Supported export formats: - CSV - JSON Error: - ERR004: Export failed
----------------------------------------------------------
[DOCS] rerank=-9.604 weighted=0.535
# Attachments (v2.3) Supported file types: - PDF - PNG - JPG Size limit: Each attachment must be 25MB or smaller. Error: - ERR002: Attachment too large
----------------------------------------------------------

!!!Potential contradictions detected:

- attachment_mb:
  • 25mb mentioned by ['docs']
  • 30mb mentioned by ['blogs']
  • 40mb mentioned by ['forums']
  • 50mb mentioned by ['forums']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(d) Q4 Output



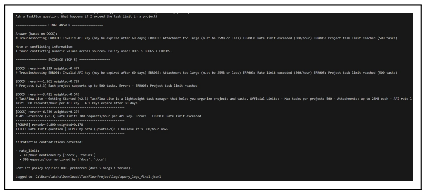
```
Ask a TaskFlow question: Can I export my tasks as a CSV file?

================ FINAL ANSWER ================

Answer (based on DOCS):
# Exporting Tasks (v2.3) Supported export formats: - CSV - JSON Error: - ERR004: Export failed

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[BLOGS] rerank=7.047 weighted=0.644
# Exporting Tasks TaskFlow Lite supports exporting to CSV and JSON. Some users request PDF export, but it is not officially supported.
----------------------------------------------------------
[DOCS] rerank=5.357 weighted=0.736
# Exporting Tasks (v2.3) Supported export formats: - CSV - JSON Error: - ERR004: Export failed
----------------------------------------------------------
[DOCS] rerank=-10.422 weighted=0.256
# TaskFlow Lite — Getting Started (v2.3) TaskFlow Lite is a lightweight task manager that helps you organize projects and tasks. Official Limits: - Max tasks per project: 500 - Attachments: up to 25MB each - API rate l
imit: 300 requests/hour per API key - API keys expire after 60 days
----------------------------------------------------------
[DOCS] rerank=-10.552 weighted=0.344
# Projects (v2.3) Each project supports up to 500 tasks. Error: - ERR005: Project task limit reached
----------------------------------------------------------
[DOCS] rerank=-11.024 weighted=0.235
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)
----------------------------------------------------------

!!!Potential contradictions detected:

- rate_limit:
  • 300/hour mentioned by ['docs']
  • 300requests/hour mentioned by ['docs']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(e) Q5 Output



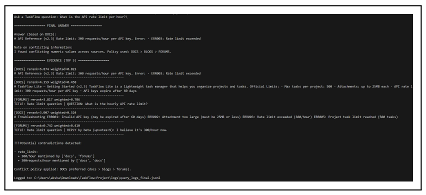
```
Ask a TaskFlow question: what is ERR003 error

================ FINAL ANSWER ================

Answer (based on DOCS):
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[DOCS] rerank=5.095 weighted=0.574
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)
----------------------------------------------------------
[DOCS] rerank=4.456 weighted=0.445
# Projects (v2.3) Each project supports up to 500 tasks. Error: - ERR005: Project task limit reached
----------------------------------------------------------
[DOCS] rerank=4.276 weighted=0.461
# API Reference (v2.3) Rate limit: 300 requests/hour per API key. Error: - ERR003: Rate limit exceeded
----------------------------------------------------------
[DOCS] rerank=3.386 weighted=0.397
# Authentication (v2.3) API keys expire after 60 days. Error Codes: - ERR001: Invalid API key
----------------------------------------------------------
[DOCS] rerank=2.631 weighted=0.346
# Exporting Tasks (v2.3) Supported export formats: - CSV - JSON Error: - ERR004: Export failed
----------------------------------------------------------

!!!Potential contradictions detected:

- rate_limit:
  • 300/hour mentioned by ['docs']
  • 300requests/hour mentioned by ['docs']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(f) Q6 Output



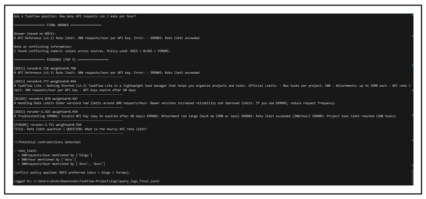
```
Ask a TaskFlow question: What happens if I exceed the task limit in a project?

================ FINAL ANSWER ================

Answer (based on DOCS):
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[DOCS] rerank=-0.339 weighted=0.477
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)
----------------------------------------------------------
[DOCS] rerank=-1.261 weighted=0.739
# Projects (v2.3) Each project supports up to 500 tasks. Error: - ERR005: Project task limit reached
----------------------------------------------------------
[DOCS] rerank=-3.421 weighted=0.545
# TaskFlow Lite — Getting Started (v2.3) TaskFlow Lite is a lightweight task manager that helps you organize projects and tasks. Official Limits: - Max tasks per project: 500 - Attachments: up to 25MB each - API rate l
imit: 300 requests/hour per API key - API keys expire after 60 days
----------------------------------------------------------
[DOCS] rerank=-6.739 weighted=0.274
# API Reference (v2.3) Rate limit: 300 requests/hour per API key. Error: - ERR003: Rate limit exceeded
----------------------------------------------------------
[FORUMS] rerank=-9.890 weighted=0.178
TITLE: Rate limit question | REPLY by beta (upvotes=9): I believe it's 300/hour now.
----------------------------------------------------------

!!!Potential contradictions detected:

- rate_limit:
  • 300/hour mentioned by ['docs', 'forums']
  • 300requests/hour mentioned by ['docs', 'docs']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(g) Q7 Output



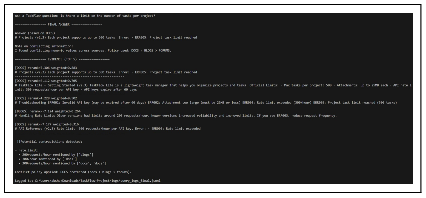
```
Ask a TaskFlow question: What is the API rate limit per hour?\

================ FINAL ANSWER ================

Answer (based on DOCS):
# API Reference (v2.3) Rate limit: 300 requests/hour per API key. Error: - ERR003: Rate limit exceeded

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[DOCS] rerank=6.874 weighted=0.823
# API Reference (v2.3) Rate limit: 300 requests/hour per API key. Error: - ERR003: Rate limit exceeded
----------------------------------------------------------
[DOCS] rerank=4.359 weighted=0.458
# TaskFlow Lite — Getting Started (v2.3) TaskFlow Lite is a lightweight task manager that helps you organize projects and tasks. Official Limits: - Max tasks per project: 500 - Attachments: up to 25MB each - API rate l
imit: 300 requests/hour per API key - API keys expire after 60 days
----------------------------------------------------------
[FORUMS] rerank=3.817 weighted=0.706
TITLE: Rate limit question | QUESTION: What is the hourly API rate limit?
----------------------------------------------------------
[DOCS] rerank=3.007 weighted=0.524
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)
----------------------------------------------------------
[FORUMS] rerank=0.742 weighted=0.410
TITLE: Rate limit question | REPLY by beta (upvotes=9): I believe it's 300/hour now.
----------------------------------------------------------

!!!Potential contradictions detected:

- rate_limit:
  • 300/hour mentioned by ['docs', 'forums']
  • 300requests/hour mentioned by ['docs', 'docs']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(h) Q8 Output



```
Ask a TaskFlow question: How many API requests can I make per hour?

================ FINAL ANSWER ================

Answer (based on DOCS):
# API Reference (v2.3) Rate limit: 300 requests/hour per API key. Error: - ERR003: Rate limit exceeded

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[DOCS] rerank=6.728 weighted=0.788
# API Reference (v2.3) Rate limit: 300 requests/hour per API key. Error: - ERR003: Rate limit exceeded
----------------------------------------------------------
[DOCS] rerank=4.777 weighted=0.494
# TaskFlow Lite — Getting Started (v2.3) TaskFlow Lite is a lightweight task manager that helps you organize projects and tasks. Official Limits: - Max tasks per project: 500 - Attachments: up to 25MB each - API rate l
imit: 300 requests/hour per API key - API keys expire after 60 days
----------------------------------------------------------
[BLOGS] rerank=1.870 weighted=0.447
# Handling Rate Limits Older versions had limits around 200 requests/hour. Newer versions increased reliability and improved limits. If you see ERR003, reduce request frequency.
----------------------------------------------------------
[DOCS] rerank=-1.425 weighted=0.454
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)
----------------------------------------------------------
[FORUMS] rerank=-3.751 weighted=0.594
TITLE: Rate limit question | QUESTION: What is the hourly API rate limit?
----------------------------------------------------------

!!!Potential contradictions detected:

- rate_limit:
  • 200requests/hour mentioned by ['blogs']
  • 300/hour mentioned by ['docs']
  • 300requests/hour mentioned by ['docs', 'docs']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(i) Q9 Output



```
Ask a TaskFlow question: Is there a limit on the number of tasks per project?

================ FINAL ANSWER ================

Answer (based on DOCS):
# Projects (v2.3) Each project supports up to 500 tasks. Error: - ERR005: Project task limit reached

Note on conflicting information:
I found conflicting numeric values across sources. Policy used: DOCS > BLOGS > FORUMS.

================ EVIDENCE (TOP 5) ================

[DOCS] rerank=7.306 weighted=0.883
# Projects (v2.3) Each project supports up to 500 tasks. Error: - ERR005: Project task limit reached
----------------------------------------------------------
[DOCS] rerank=6.112 weighted=0.705
# TaskFlow Lite — Getting Started (v2.3) TaskFlow Lite is a lightweight task manager that helps you organize projects and tasks. Official Limits: - Max tasks per project: 500 - Attachments: up to 25MB each - API rate l
imit: 300 requests/hour per API key - API keys expire after 60 days
----------------------------------------------------------
[DOCS] rerank=4.228 weighted=0.502
# Troubleshooting ERR001: Invalid API key (may be expired after 60 days) ERR002: Attachment too large (must be 25MB or less) ERR003: Rate limit exceeded (300/hour) ERR005: Project task limit reached (500 tasks)
----------------------------------------------------------
[BLOGS] rerank=-7.124 weighted=0.264
# Handling Rate Limits Older versions had limits around 200 requests/hour. Newer versions increased reliability and improved limits. If you see ERR003, reduce request frequency.
----------------------------------------------------------
[DOCS] rerank=-7.177 weighted=0.316
# API Reference (v2.3) Rate limit: 300 requests/hour per API key. Error: - ERR003: Rate limit exceeded
----------------------------------------------------------

!!!Potential contradictions detected:

- rate_limit:
  • 200requests/hour mentioned by ['blogs']
  • 300/hour mentioned by ['docs']
  • 300requests/hour mentioned by ['docs', 'docs']

Conflict policy applied: DOCS preferred (docs > blogs > forums).

Logged to: C:\Users\aksha\Downloads\TaskFlow-Project\logs\query_logs_final.jsonl
```

(j) Q10 Output

Figure 1: Placeholder grid for 10 screenshot outputs (replace images in `figs/`).

11 Limitations and Future Work

- The contradiction detector focuses on numeric claims (MB/days/hour). More general contradiction detection would require deeper semantic modeling.
- Hit@1 decreased after reranking in this run :contentReference[oaicite:17]index=17; a refinement is to incorporate source priors directly into reranking or to select the highest reranked chunk within the preferred source.

- The answer synthesis is deterministic (no external LLM). Adding a small LLM could improve readability, but would introduce nondeterminism and potential hallucinations.