

Evaluation Criteria: RAG Challenge

1. Retrieval & Answer Quality (Core Metric)

This is the primary metric, evaluating how well the system answers predefined test questions.

Scoring (Each 0-2 points per question):

- **Relevance:** Does the answer match the intent of the question? *(Evaluated using LLM-based testing)*
- **Grounding:** Is the answer clearly based on the provided HTML data? *(Compared with ground-truth answers)*
- **Language Quality:** Is the response well-structured, grammatically correct, and coherent? *(LLM-based testing)*
- **Efficiency & Runtime:** Is the system efficient in terms of retrieval and response generation? *(Measured as execution time)*

👉 **Weighting:** High impact on the overall score.

2. System Architecture & Retrieval Logic

Assesses the design choices and implementation of retrieval methods. Make sure that you cover main points in the presentation.

Scoring (Each 1-5 points):

- **Code Organization & Project Structure:** Is the code well-structured, modular, and easy to run?
- **Data Processing & Storage Design:** Is the choice of database and processing pipeline logical and efficient? *(Evaluated via human analysis)*
- **Retrieval Approach:** Are indexing, chunking, and search methods effectively implemented?

👉 **Weighting:** Medium-High, judged based on presentation + technical review.

3. Bonus / Creativity

Encourages innovation, extra effort, and usability enhancements.

Scoring:

- **Bonus Features (0-3 points):** Additional functionality like CLI, UI, dashboards, or visualizations.
- **Novelty / Creativity (0-2 points):** Unique approaches, hybrid methods, or novel integrations.
- **Code Clarity & Usability (Each 1-5 points):**
 - **Clarity & Structure:** Is the code modular, documented, and easy to understand?
 - **Setup & Runability:** Can the solution be executed easily with a README or notebook?
- **API Call Costs:** Is the system optimized to minimize API calls and cost? (*Tracked in the code logs*)

👉 **Weighting: Low-Medium**, used to break ties and reward ambitious solutions.

Final Scoring:

- **Retrieval & Answer Quality** → ~50-60%
- **System Architecture & Retrieval Logic** → ~30-40%
- **Bonus / Creativity** → ~10-20%

Participants are encouraged to optimize their solutions for **accuracy, efficiency, and usability** while also exploring creative enhancements!

3. Example questions:

1. What company provides assisted living near Richmond, Virginia?
 - a. covenantwoods.com (and others possible)
2. What companies uses packaging materials in Valencia, California?
 - a. amsfulfillment.com (and others possible)
3. Who uses Agile Methodologies to deal with Marketing in Fort Lauderdale, FL?
 - a. starmark.com (and others possible)