

SUPPLEMENTARY DATASET FOR CULTURAL TOURISM CHATBOT EVALUATION

Dataset Overview

This dataset provides a structured evaluation record of a Retrieval-Augmented Generation (RAG) system designed for disseminating cultural tourism information. It focuses on assessing response faithfulness, contextual grounding, and hallucination risk through a combined automatic and Human-in-the-Loop (HITL) evaluation framework.

This documentation is intended to be read in conjunction with the primary research article, which provides the theoretical foundation and relevant citations, related work, and an in-depth analysis of the results. Here, we focus on the full specifications and composition of the dataset itself.

Each dataset entry represents a single evaluation unit and includes the original user query, retrieved knowledge chunks, generated responses, and multiple layers of assessment outputs, including independent judgments from three human evaluators and a reconciled final decision.

Automated evaluation metrics are computed to provide baseline signals of semantic similarity and contextual relevance, while LLM-based judgments offer preliminary qualitative assessments. Final validation is conducted through structured HITL evaluation to ensure interpretability and consistency in decision governance.

The dataset emphasizes traceability and reproducibility by preserving all intermediate artifacts required to reconstruct the evaluation process. Error categories and decision rationales are explicitly documented to distinguish retrieval failures, generation errors, and cases of context misuse. By integrating automatic and human-centered evaluation stages, the dataset supports methodological analysis of RAG system reliability and evaluation alignment. It is intended for use as a reference dataset in research on RAG evaluation, HITL methodologies, and faithfulness assessment in domain-specific generative AI systems.

TABLE OF CONTENT

Dataset Overview.....	i
1. Introduction	1
2. Data Sources: Curated Cultural Documents	3
3. Chunking and Knowledge Unit Construction	4
4. Query Design and Experimental Inputs.....	6
5. LLM Orchestration: Candidate and Reference Generation	7
6. Automatic Screening Layer	9
7. Human-in-the-Loop (HITL) Evaluation and Decision Policy.....	11
7.1. HITL Evaluation Artifacts and Annotation Schema	12
7.2. Separation of System-Level Failure and Content-Level Evaluation	12
7.3. Multi-Evaluator HITL Decision Process.....	12
7.4. Decision Outcomes and Error Code Framework.....	13
7.5. Handling Asymmetric Candidate-Reference Conditions	14
7.6. Inter-Evaluator Reliability and Quantitative Overview.....	14
7.7. Example of HITL Annotation Representation	14
7.8. Role of HITL in the Final Dataset	16
8. Dataset Governance, Traceability, and Reusability.....	16
8.1. Dataset Governance	16
8.2. Traceability and Auditability	16
8.3. Dataset Versioning and Reproducibility	17
8.4. Reusability and Scope of Use.....	17
8.5. Ethical Considerations and Limitations.....	17
9. Conclusion and Future Work.....	19
Appendix A: Overview of Curated Cultural Documents (Original Indonesian)	20
Appendix B: Chunk Schema	24
Appendix C: The Completed Query Categories and Texts.....	30
Appendix D: Candidate and Reference Outputs (Original Indonesian).....	33
Appendix E: The Completed Automatic Metric Schema	44
Appendix F: Human-in-the-Loop (HITL) Annotation Schema.....	47

Appendix G: Distribution of Evaluation Scores per Evaluator	48
Appendix H : Pairwise Inter-Evaluator Agreement Based on Cohen’s Kappa	49
Appendix I: Multi-Evaluator Agreement Using Fleiss’ Kappa	50
Appendix J: The Full Set HITL Evaluation Results	51

FIGURE

Figure 1. HITL-RAG Evaluation Framework.	2
Figure 2. Chunk Generation for RAG Memory Injection.....	5
Figure 3. Automatic Screening Layer with Metric and Thresholds.....	10
Figure 4. HITL Evaluation Artifact and Annotation Schema.	11

TABLE

Table 1. Metadata Schema and Traceability Attributes.....	3
Table 2. Overview of Curated Cultural Documents	4
Table 3. Chunk Schema	5
Table 4. Query Categories and Text Examples	6
Table 5. Candidate and Reference Output Examples.....	8
Table 6. Automatic Metric Schema	10
Table 7. Human-in-the-Loop Annotation Schema is attached in Appendix F.	12
Table 8. HITL Error Code Definition (E0-E10).....	13
Table 9. Example of HITL Annotation Record (Excerpt).....	15

1. Introduction

Retrieval-Augmented Generation (RAG) has been widely adopted as an approach to improve the factual grounding of large language models by incorporating external knowledge sources into the generation process. By retrieving relevant textual evidence prior to response generation, RAG systems aim to mitigate hallucination and enhance response reliability, particularly in knowledge-intensive application domains.

However, the integration of retrieval does not fully eliminate generation errors. RAG systems may still exhibit failure modes such as irrelevant retrieval, incomplete grounding, or the inappropriate use of retrieved context. These issues are especially critical in domains where informational accuracy and cultural sensitivity are essential, including cultural tourism information. In such settings, inaccurate or misleading responses may distort cultural narratives and reduce user trust.

Assessing the faithfulness of RAG-generated responses therefore remains a methodological challenge. Automatic evaluation metrics provide scalable and reproducible signals but are often insufficient to capture nuanced grounding errors or contextual misuse. In contrast, purely human-based evaluation offers interpretive depth but is difficult to standardize and resource-intensive when applied at scale. This gap has motivated the development of Human-in-the-Loop (HITL) evaluation frameworks that integrate automated assessment with structured human judgment.

This dataset was developed to support systematic analysis of faithfulness and grounding in RAG systems through a controlled HITL evaluation process. It records the complete evaluation pipeline, including user queries, retrieved knowledge chunks, generated responses, automated evaluation outputs, and human validation results. Emphasis is placed on preserving traceability across evaluation stages, enabling detailed examination of how retrieval and generation behaviors influence final evaluation outcomes.

The dataset focuses on cultural tourism information derived from narrative and descriptive cultural sources. This domain poses distinct challenges due to its reliance on contextual interpretation, implicit meanings, and culturally embedded references. As a result, it provides a suitable testbed for examining the alignment between retrieved knowledge, generated responses, and human evaluative reasoning.

Figure 1 illustrates the conceptual architecture of the HITL-based evaluation framework used in this study. This design explicitly separates functional roles throughout the evaluation flow, including candidate response generation, reference construction and source-based chunk retrieval by the LLM Judge, and automated filtering through quantitative metrics, culminating in final validation by a human evaluator.

This framework operates as a methodological contract to define decision-making authority and governance boundaries within the workflow. Automated metrics and model-based assessments are positioned as supporting signals to assist the human layer, rather than as the final arbiter. Thus, the Human-in-the-Loop component retains ultimate authority for accuracy assessment, ensuring that no single automated module can independently determine the evaluation outcome. By implementing this phased and segregated process based on roles, the framework effectively reduces systemic evaluation bias and increases traceability across all stages of the assessment.

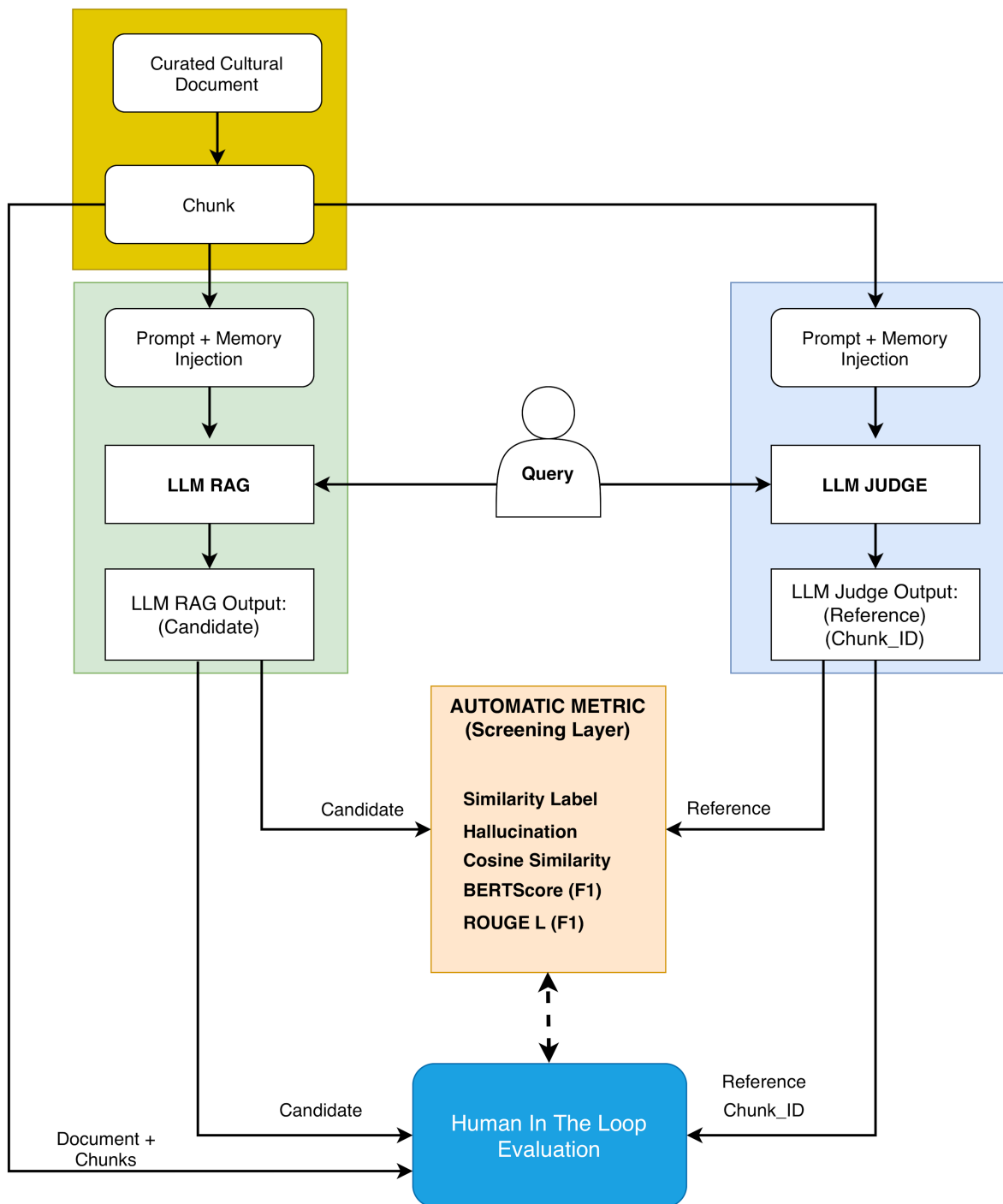


Figure 1. HITL-RAG Evaluation Framework.

Overview of the Human-in-the-Loop (HITL) architecture for assessing faithfulness and grounding in Retrieval-Augmented Generation (RAG) systems. The diagram delineates the functional separation between retrieval, generation, automated screening, and human validation phases.

To operationalize the governance framework illustrated in Figure 1, a structured metadata schema is implemented to ensure transparency and end-to-end traceability. This schema serves as a bridge, systematically connecting raw cultural data, automated system traces, and human adjudication

outcomes. The detailed metadata attributes and their functional roles in ensuring data faithfulness are presented in Table 1.

Table 1. Metadata Schema and Traceability Attributes

Category	Metadata Field	Description	Functional Role
Provenance	Query_ID & Chunk_ID	Unique identifiers for user queries and their corresponding reference text units.	Ensures absolute traceability from the query back to the grounding source.
Technical Trace	Automated Metrics & Threshold	Scores for BERTScore (F1), ROUGE-L (F1), Cosine Similarity, Similarity Label, Hallucination	Provides a quantitative and qualitative baseline for the initial system performance.
Governance	Error Taxonomy	Error classification using codes E0 through E10.	Standardizes the identification of hallucination types and context misuse.
Final Result	Human Decision	Final status: <i>Accepted</i> , <i>Rejected</i> , or <i>Abstain</i> .	Acts as the gold standard for validating or overriding machine-based judgments.
Qualitative	Rationalization	Textual justification provided by human evaluators.	Offers deep insights into the reasoning behind the final adjudication.

With the metadata structure established, the following sections detail the data lifecycle sequentially, starting from the curated cultural document sources in Section 2 to the technical validation of the final evaluation records.

2. Data Sources: Curated Cultural Documents

This dataset is derived from a collection of curated cultural documents that were manually compiled and selected by the researchers as the primary knowledge source for the evaluation process. These documents contain cultural narratives, site descriptions, and historical context relevant to the cultural tourism chatbot domain that constitutes the focus of this study.

All documents are stored in Google Document format and were not directly sourced from open corpora. This manual curation approach was adopted to ensure narrative consistency, clarity of cultural context, and controlled provenance of the reference sources used in the evaluation. Accordingly, the cultural documents function as a conceptually accountable knowledge base rather than as an unstructured collection of raw text.

Within the context of this dataset, the cultural documents are not used directly by the model. Instead, they are first processed through a chunking stage to construct structured knowledge units that serve as explicit grounding references throughout all evaluation stages. In this design, the cultural documents are positioned as upstream knowledge sources, while the resulting chunks operate as the practical representation of ground truth at the evaluation level.

As an initial overview, Table 2 summarizes the cultural documents used in this dataset, including document identifiers, thematic domains, and approximate text length. This dataset version utilizes a single primary cultural narrative (Ande-Ande Lumut) as a focused case study. The complete list of documents, along with detailed metadata, is provided separately in the Appendix.

Table 2. Overview of Curated Cultural Documents

Document_ID	Title	Domain	Length (approx.)
Doc_01	Cerita Budaya Ande Ande Lumut	Cultural Stories	1,250 words

(Full documents and detailed metadata are available in Appendix A.)

3. Chunking and Knowledge Unit Construction

The curated cultural documents described in the previous section are subsequently processed through a chunking stage to construct structured knowledge units that can be used operationally in the evaluation process. This stage is executed using a chunk generator orchestrated through an n8n workflow, ensuring that document segmentation is performed in a consistent and well-documented manner.

The chunking process divides each cultural document into text segments of controlled size, referred to as chunks. Each chunk is assigned a unique identifier (`chunk_id`) along with positional information relative to the source document. Through this approach, each chunk is treated as an atomic reference unit, enabling direct linkage between model responses, evaluation outcomes, and the underlying knowledge source.

In this dataset, chunks function as the operational representation of source-based ground truth. All response generation processes, including those performed by the LLM RAG and the LLM Judge, as well as all evaluation stages, explicitly refer to specific chunk identifiers. This design ensures that every evaluation decision can be traced back to a relevant source text segment rather than to a document as a whole.

Importantly, chunk texts are preserved in their original Indonesian language and are not translated into English. This decision is made to maintain source fidelity, semantic integrity, and traceability throughout the evaluation process. Translating the chunk content would alter the grounding reference and introduce an additional transformation layer that is not part of the evaluated system behavior. Consequently, only the methodological descriptions, field labels, and annotation metadata are presented in English, while all ground truth content remains unchanged.

Figure 2 illustrates the process of constructing chunks from cultural documents and their subsequent use as knowledge memories injected into the LLM RAG and LLM Judge systems. Although memory injection is implemented at a later stage, the chunk structure generated in this step forms the basis of the entire evaluation pipeline. The figure visualizes the transformation of curated cultural documents into chunk-level basic units that remain in their original Indonesian

form to maintain source fidelity. These chunks serve as atomic ground truth references throughout the evaluation process.

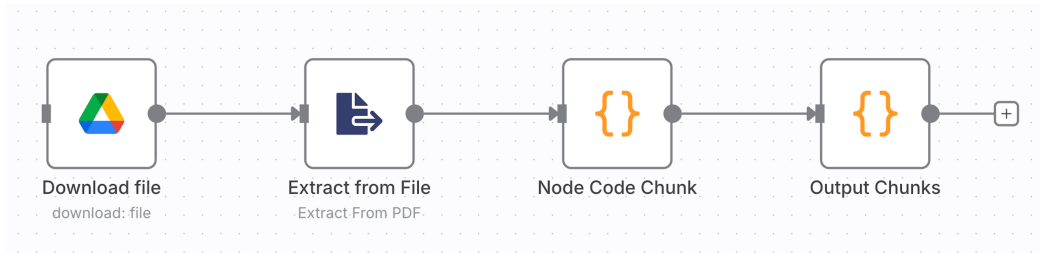


Figure 2. Chunk Generation for RAG Memory Injection.

Process of converting curated cultural documents into Indonesian-language text chunks that serve as atomic grounding units in the RAG system’s knowledge memory.

As an illustration of the data structure, Table 3 presents the chunk schema used in the dataset, including the relationship between the source document, positional index, chunk identifier, and chunk text. The complete version of this table, including all chunks used in the experiment, is provided in the Appendix.

Table 3. Chunk Schema

Document_ID	Index	Chunk_ID	Chunk_Text (Original Indonesian Source)
Doc_01	1	0	Bagian paragraf ini disusun sebagai penjelasan konseptual untuk menggambarkan potensi kesenjangan ...
	2	500	penting dalam evaluasi human-judge atau penilaian semantik oleh manusia untuk memastikan ...
	3	1000	karya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa ...
	4	1500	Pada intinya, Ande-Ande Lumut adalah epos tentang pencarian jodoh seorang ...
	5	2000	Namun, Ande-Ande Lumut menolak semua tawaran itu, seolah ia tengah mencari ...
...
	17	7500	yang magis, menciptakan atmosfer yang memukau dan pengalaman budaya yang ...
	18	8000	Platform digital seperti YouTube dan aplikasi dongeng juga menjadi ...
	19	8500	dalam bentuk ketoprak atau wayang orang, seringkali dengan sentuhan kontemporer ...
	20	9000	yang menghubungkan masa lalu, kini, dan masa depan. Ia adalah penjaga nilai-nilai ...
	21	9500	Additional chunks for abstention and failure trapping ...

(The complete schema and full set of chunks are available in Appendix B.)

Note on Ground Truth Scope: Chunks with IDs chunk-0 and chunk-500 contain methodological explanations, and chunk-9500 is a system fallback chunk. These are not used as factual grounding references for user queries. All content-level faithfulness assessments are conducted exclusively

with respect to chunks representing curated cultural knowledge. See Appendix B for a detailed clarification.

4. Query Design and Experimental Inputs

This dataset employs a systematically designed set of queries to examine the behavior of a retrieval augmented generation (RAG) system under diverse knowledge context conditions. Queries serve as the primary trigger for the construction of all dataset artifacts, as every data unit, ranging from model responses to evaluation outcomes, is derived from the same query instance.

Each query is classified into one of three contextual categories:

- In-context, referring to questions whose answers are explicitly available in one or more source chunks.
- Ambiguous, referring to questions that allow multiple interpretations or require contextual reasoning across chunks.
- Out-of-context, referring to questions that are not supported by any available chunk and are intentionally designed to assess the system’s ability to recognize knowledge limitations and avoid hallucinated responses.

This classification scheme ensures that the dataset evaluates not only the system’s ability to answer factual questions, but also its behavior when confronted with uncertainty or the absence of supporting information.

In particular, out-of-context queries are intentionally designed to elicit epistemic abstention behavior, allowing the dataset to capture cases where a faithful response consists of explicitly acknowledging knowledge limitations rather than generating speculative content.

Each query is synchronously provided to two model pathways, namely the LLM RAG and the LLM Judge, using the same query identifier (`query_id`). The query is preserved within the workflow context throughout execution, allowing all outputs, including candidate responses, reference outputs, and evaluation artifacts, to be aligned within a single dataset record.

As an illustration, Table 4 presents representative examples of queries used in the dataset along with their contextual categories. The complete list of queries, including question variants and category distribution, is provided in the Appendix.

Table 4. Query Categories and Text Examples

Query_ID	Category	Query_Text (Original Indonesian)
Q1	In-context	Mengapa Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau?
Q2	In-context	Bagaimana sikap Kleting Kuning sepanjang perjalanan?
...
Q6	Ambiguous	Apa makna kisah Ande-Ande Lumut bagi kehidupan manusia?
Q9	Ambiguous	Apa pesan utama perjalanan Kleting Kuning?
...

Query_ID	Category	Query_Text (Original Indonesian)
Q26	Out-of-context	Siapa pengarang asli cerita Ande-Ande Lumut menurut dokumen?
Q56	Out-of-context	Sebutkan tahun dan lokasi ditemukannya naskah kuno Ande-Ande Lumut yang menjadi rujukan utama cerita ini.
Q57	Out-of-context	Bagaimana hubungan genealogis Ande-Ande Lumut dengan tokoh-tokoh kerajaan Jawa lainnya menurut catatan sejarah?

(The complete set of queries is available in Appendix C.)

5. LLM Orchestration: Candidate and Reference Generation

For each query described in the previous section, the dataset executes two functionally distinct response generation processes, namely through the LLM RAG and the LLM Judge. Both processes operate on the same knowledge source in the form of retrieved chunks, but they are designed to produce different artifacts in accordance with their respective evaluation roles.

The LLM RAG functions as the candidate generator, producing system responses to each query by leveraging injected chunks within its context window as a form of knowledge memory. The output generated at this stage is recorded as the candidate response and represents the observable behavior of the RAG system under evaluation. The LLM Judge does not function as a normative evaluator but serves as a consistency-checking instrument whose outputs are fully subordinate to human judgment.

In contrast, the LLM Judge functions as a reference and attribution generator. This model receives the same query and the same set of chunks, but operates under a different prompt configuration that instructs it to generate a source-based reference response and to identify the chunk identifiers considered most relevant for grounding. Accordingly, the reference output is not positioned as an evaluation of the candidate response, but as an independently generated, source-grounded answer.

Although both models are executed within the same workflow, the LLM RAG and the LLM Judge operate in parallel at the logical level. Each model receives the query and chunks independently. The sequential execution order observed in the workflow does not imply semantic dependency between models, but reflects an orchestration design that facilitates record-level data alignment.

As illustrated in Figure 3, the n8n workflow maintains a shared execution context, ensuring the original query remains accessible to both the LLM RAG and LLM Judge nodes throughout the pipeline. It should be noted that the input to LLM Judge is obtained directly from the query stored in the workflow context, not from the LLM RAG output. This architecture ensures that the query, candidate response, reference output, and chunk identifier are recorded in a single dataset entry without introducing any logical dependencies. Although the workflows are executed sequentially for orchestration, they do not imply semantic dependencies; instead, both nodes operate independently on the same data. The n8n workflow functions strictly as a synchronization mechanism to maintain proper record-level alignment.

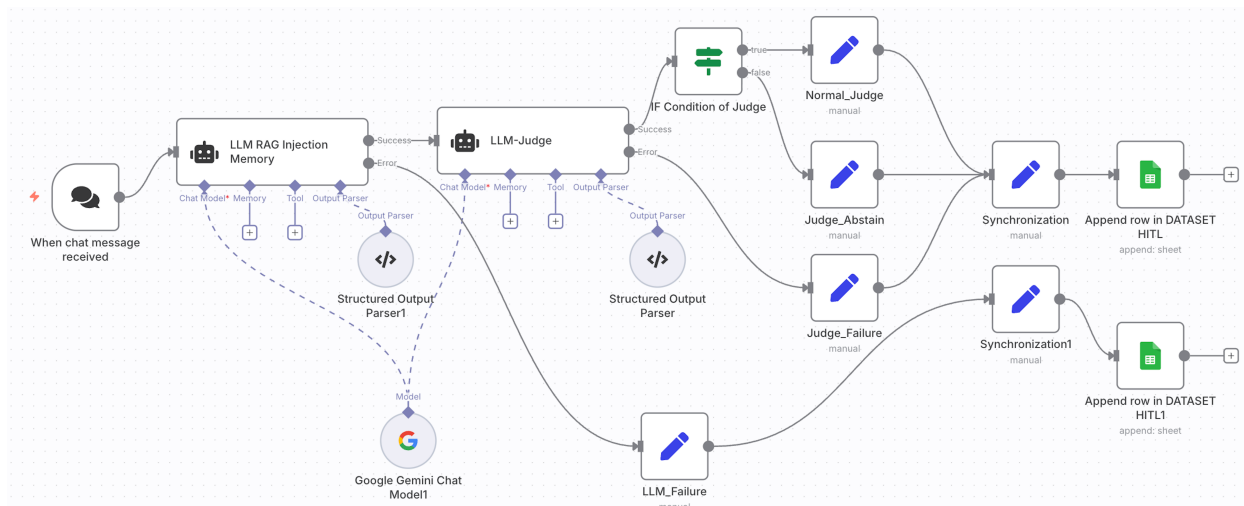


Figure 3. n8n Workflow Orchestration for Candidate, Reference Generation, and Chunk_ID Retrieval.

Illustration of the parallel execution of LLM RAG (candidate generator) and LLM Judge (reference generator, chunk_id retrieval) within the same query context, ensuring record-level alignment without semantic dependency.

As an illustration, Table 5 presents representative examples of candidate-reference pairs along with the corresponding chunk identifiers generated from the same query. The complete version of this table, including all queries and model outputs, is provided in the Appendix.

Table 5. Candidate and Reference Output Examples

Query_ID	Candidate_Text (Original Indonesian)	Reference_Text (Original Indonesian)	Chunk_ID
Q1	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena mereka memiliki aroma amis Yuyu Kangkang yang melekat pada diri mereka. Ini menunjukkan bahwa mereka telah mengorbankan martabat demi kemudahan menyeberangi sungai, berbeda dengan Kleting Kuning yang berintegritas.	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena aroma amis Yuyu Kangkang yang melekat pada diri mereka.	chunk-3000
...
Q6	Kisah Ande-Ande Lumut mengajarkan bahwa budi pekerti, kesabaran, dan integritas jauh lebih berharga daripada penampilan fisik atau kekayaan. Cerita ini menekankan pentingnya karakter, kejujuran, dan kebaikan hati dalam menghadapi cobaan hidup, serta keberanian untuk memegang teguh prinsip moral.	Kisah Ande-Ande Lumut mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran, ketabahan, keberanian, dan integritas dalam menghadapi cobaan, serta memegang teguh prinsip kebenaran dan kejujuran.	chunk-4000, chunk-4500, chunk-5000, chunk-5500
...

Query_ID	Candidate_Text (Original Indonesian)	Reference_Text (Original Indonesian)	Chunk_ID
Q27	Informasi mengenai tahun pertama kali cerita Ande-Ande Lumut ditulis tidak tersedia dalam dokumen yang diberikan. Dokumen tersebut menjelaskan bahwa Ande-Ande Lumut adalah warisan budaya takbenda yang dituturkan secara lisan dan telah melampaui batasan waktu.	Maaf saya belum mendapatkan pelatihan tentang pertanyaan yang anda berikan, silahkan ajukan pertanyaan lainnya.	chunk-9500
...
Q60	Para Kleting tidak menyeberang sungai menggunakan speedboat karena cerita Ande-Ande Lumut adalah dongeng tradisional yang berlatar waktu sebelum teknologi modern seperti speedboat ada. Penyeberangan sungai dalam cerita tersebut melibatkan sosok mitologis Yuyu Kangkang.	Para Kleting menyeberang sungai yang dijaga Yuyu Kangkang. Kleting Kuning menyeberang dengan kekuatannya sendiri.	chunk-2500, chunk-3000
...

(The complete table is available in Appendix D.)

6. Automatic Screening Layer

After candidate responses and reference outputs are generated from the same query, the dataset applies an automatic screening layer as an initial evaluation stage based on quantitative metrics. This stage is intended to provide an early indication of the degree of alignment between the RAG system's response and the source-based reference response, prior to human evaluation.

The automatic screening process is executed using a separate Python script that is integrated within the n8n workflow. Input data for this stage are taken directly from the aligned outputs of the LLM RAG (candidate) and the LLM Judge (reference) at the record level. Through this approach, each metric value can be explicitly traced back to the same query, candidate response, reference response, and associated chunk identifiers.

The automatic evaluation metrics used in this dataset include:

- BERTScore (F1), to measure semantic similarity based on contextual representations,
- ROUGE-L(F1), to capture similarity based on sequence overlap,
- Cosine Similarity, to measure vector-based proximity between responses.

In addition to numeric values, the system produces derived outputs in the form of similarity labels, such as high, medium, or low, as well as preliminary indications of potential hallucination. These labels and indicators are not intended to function as final decisions. Instead, they serve as supporting signals to assist human evaluators in the subsequent Human-in-the-Loop (HITL) stage.

Automatic evaluation metrics are computed for all candidate-reference pairs successfully generated by the system under logically valid conditions. Metric computation is independent of

the final HITL decision, whether acceptance, abstention, or failure at the content level. Metrics are not computed for system-level failures in which no valid response is available for analysis. In this context, automatic metrics function as diagnostic signals that complement, but do not replace, human judgment.

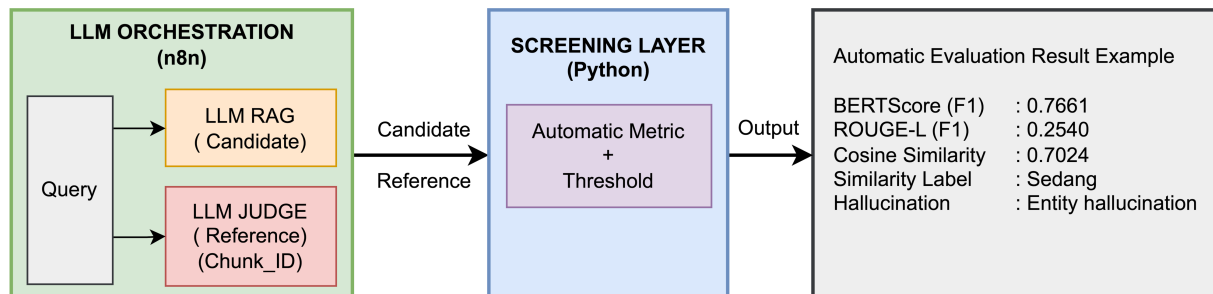


Figure 3. Automatic Screening Layer with Metric and Thresholds.

The depiction of the automatic evaluation stage using BERTScore (F1), ROUGE-L (F1), Cosine Similarity equipped with label similarity threshold and hallucination to provide diagnostic signals for further HITL evaluation.

As illustrated in **Figure 4**, the automatic screening layer serves as a specialized filtering stage within a multi-stage judgment framework for all logically valid outputs. In this architecture, the diagnostic thresholds shown are utilized exclusively for screening purposes; they do not carry decision-making authority nor do they determine the final acceptance or rejection of a response. Instead, the results from this layer are integrated as supplemental contextual metadata to inform and enhance the Human-in-the-Loop (HITL) evaluation process. By positioning these automated insights as diagnostic tools rather than final arbiters, the framework ensures that human evaluators remain the primary authority while benefiting from the precision of automated screening.

As an illustration of the data structure, Table 6 presents the automatic metric schema recorded in the dataset, showing the relationship between the query identifier and the computed metric values. The complete version of this table is provided in the Appendix.

Table 6. Automatic Metric Schema

Query_ID	Automatic Metric Data (BERTScore (F1), ROUGE-L (F1), Cosine Similarity, Similarity_Label, Hallucination)
Q1	{'BERTScore (F1)': 0.8878, 'ROUGE-L (F1)': 0.6667, 'Cosine_Similarity': np.float32(0.9415), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q2	{'BERTScore (F1)': 0.8335, 'ROUGE-L (F1)': 0.4848, 'Cosine_Similarity': np.float32(0.8809), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q3	{'BERTScore (F1)': 0.7661, 'ROUGE-L (F1)': 0.2540, 'Cosine_Similarity': np.float32(0.7024), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q4	{'BERTScore (F1)': 0.8050, 'ROUGE-L (F1)': 0.4375, 'Cosine_Similarity': np.float32(0.9036), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Entity hallucination'}
Q5	{'BERTScore (F1)': 0.8247, 'ROUGE-L (F1)': 0.4615, 'Cosine_Similarity': np.float32(0.8733), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Entity hallucination'}

Query_ID	Automatic Metric Data (BERTScore (F1), ROUGE-L (F1), Cosine Similarity, Similarity_Label, Hallucination)
...	...
Q60	{'BERTScore (F1)': 0.7551, 'ROUGE-L (F1)': 0.2917, 'Cosine_Similarity': np.float32(0.6295), 'Similarity_Label': 'Rendah', 'Hallucination': 'Entity hallucination'}
Q61	{'BERTScore (F1)': 0.7620, 'ROUGE-L (F1)': 0.2553, 'Cosine_Similarity': np.float32(0.8210), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}

(The complete metric data are available in Appendix E.)

7. Human-in-the-Loop (HITL) Evaluation and Decision Policy

Human-in-the-Loop (HITL) evaluation constitutes the final and decisive validation layer in the construction of this dataset. At this stage, human evaluators act as the ultimate decision authority by assessing the faithfulness of system-generated responses with respect to the user query and the retrieved knowledge chunks used for grounding.

Unlike automated evaluation, HITL does not aim to optimize surface-level textual similarity. Instead, it focuses on determining whether generated responses are relevant to the query, consistent with the referenced knowledge sources, and free from unsupported or unverifiable claims.

All artifacts produced in the preceding stages, including candidate responses, reference responses, chunk identifiers, and automatic screening metrics, are aligned at the record level and jointly presented to support structured and traceable human evaluation. To operationalize this human-centered validation process, HITL evaluation relies on a structured set of aligned artifacts that collectively represent the query, system outputs, and auxiliary assessment signals.

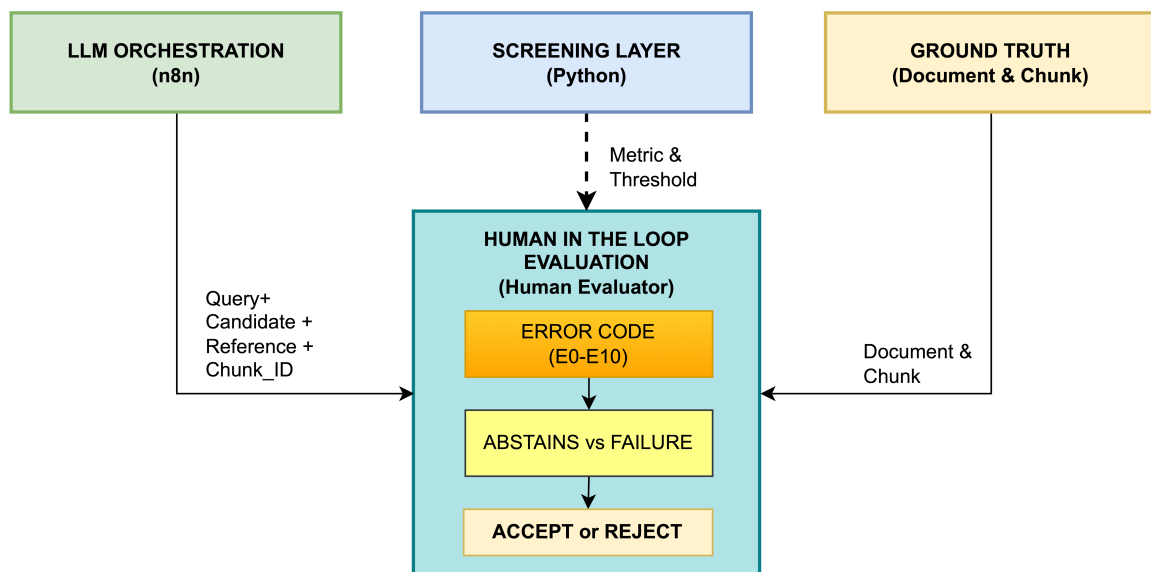


Figure 4. HITL Evaluation Artifact and Annotation Schema.

A structured package presented to a human evaluator contains candidate responses, reference responses, section identifiers, automated metrics, and curated culture documents along with chunks for final fidelity assessment.

As illustrated in Figure 5, this is the final stage in the multi-stage assessment framework. All artifacts managed within all components of the framework are an essential part of the HITL evaluation process.

7.1. HITL Evaluation Artifacts and Annotation Schema

For each query-response pair, evaluators are provided with a structured HITL evaluation package consisting of:

- the candidate response generated by the RAG system,
- the reference response and associated chunk identifier produced by the LLM Judge, and
- auxiliary automatic screening metrics, such as similarity scores and hallucination indicators.

All artifacts are synchronized at the record level through an automated workflow, enabling full traceability between the query, system outputs, and human judgments. The complete field-level annotation schema used in the HITL process is provided separately to ensure clarity and reproducibility.

Table 7. Human-in-the-Loop Annotation Schema is attached in Appendix F.

Based on this unified artifact package, the HITL process applies a decision policy that explicitly separates technical system failures from content-level evaluation outcomes.

7.2. Separation of System-Level Failure and Content-Level Evaluation

The HITL decision policy explicitly distinguishes between system-level failures and content-level evaluation errors. This separation is governed by the logical status output of the LLM Judge.

a) System-Level Failure

When the LLM Judge fails to produce a valid reference response or chunk attribution, the record is classified as a system-level failure. In such cases, grounding cannot be verified, and content-level evaluation is not performed. All records under this condition are directly labeled as system failures.

b) Content-Level Evaluation

When a valid reference response and chunk identifier are available, the record proceeds to content-level evaluation. At this stage, potential inaccuracies, weak grounding, or hallucinations are treated as content-related issues rather than system failures and are fully evaluated by human annotators.

Once records enter the content-level evaluation pathway, assessment responsibility is transferred entirely to human evaluators operating under a multi-evaluator HITL framework.

7.3. Multi-Evaluator HITL Decision Process

Each query-response pair is independently evaluated by three human evaluators using a shared rubric. Evaluators assess responses without enforced role specialization, allowing diverse but complementary perspectives to emerge naturally.

The final HITL decision recorded in the dataset reflects a reconciliation of the three independent judgments. Disagreements are resolved through majority agreement, with predefined priority rules applied for abstentions and system failures. Outcomes from this multi-evaluator assessment are formalized through a predefined set of decision labels and error codes to ensure consistency and interpretability across the dataset.

7.4. Decision Outcomes and Error Code Framework

For content-level evaluation, each record is assigned a single primary HITL decision and a corresponding error code based on predefined criteria. Decision outcomes include acceptance, abstention, or failure.

To support consistent and granular assessment, the dataset employs an eleven-level error code framework (E0-E10), distinguishing fully grounded responses, varying degrees of content deficiencies, abstentions, and system-level failures.

Table 8. HITL Error Code Definition (E0-E10)

Error Code	Evaluation Category	Description	Failure Type
E0	Fully Grounded	Response is fully relevant, correctly answers the query, and is explicitly supported by the referenced chunk.	-
E1	Minor Omission	Response is correct and source-supported, but omits minor details that do not alter the main meaning.	-
E2	Partial Grounding	Response is partially supported by the chunk but includes additional information not fully traceable to the source.	Content
E3	Weak Attribution	Response is relevant, but the referenced chunk supports only a limited portion of the content.	Content
E4	Context Drift	Response deviates from the focus of the query while remaining within the same general domain.	Content
E5	Ambiguous Response	Response is overly general or ambiguous, preventing clear verification against the source.	Content
E6	Unsupported Claim	Response contains factual claims not supported by any chunk.	Content
E7	Hallucination	Response appears linguistically plausible but lacks valid source support in the chunks.	Content
E8	Contradictory Content	Response directly contradicts information contained in the source chunk.	Content
E9	Abstain / No-Answer	Response explicitly declines to answer or avoids providing information.	-
E10	System Failure	System fails to produce an evaluable output, such as empty output or processing error.	System

In addition to standard evaluation outcomes, the HITL framework explicitly addresses asymmetric conditions between candidate and reference responses that require special handling.

7.5. Handling Asymmetric Candidate-Reference Conditions

The dataset defines explicit rules for cases in which candidate responses and reference responses are asymmetric. These rules ensure that responses are never evaluated without a verifiable grounding basis and that technical failures are not conflated with content-level errors.

7.6. Inter-Evaluator Reliability and Quantitative Overview

To assess the consistency of the multi-evaluator HITL process, inter-rater agreement statistics were computed based on independent evaluations prior to consensus. These statistics serve as supporting evidence for annotation reliability rather than as primary analytical targets.

Quantitative summaries include evaluator score distributions and standard agreement coefficients for both pairwise and multi-rater scenarios. These results are provided in Appendix G for score distributions, Appendix H for pairwise Cohen's Kappa, and Appendix I for Fleiss' Kappa.

7.7. Example of HITL Annotation Representation

To illustrate the application of the HITL framework, a representative excerpt of annotated records is presented. This example demonstrates how evaluator scores, consensus decisions, and qualitative notes are recorded for individual query-response pairs.

Table 9. Example of HITL Annotation Record (Excerpt)

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query	Candidate_Response
Q1	E0	E4	E2	Accepted	The core answer is supported by the chunk (the foul smell of Yuyu Kangkang), but there is an addition of a normative interpretation about sacrificing dignity and integrity that is not explicitly stated in chunk-3000.	No Discussion Needed	Mengapa Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau?	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena mereka memiliki aroma amis Yuyu Kangkang yang melekat pada diri mereka. [...]
Q11	E9	E0	E9	Accepted	The answer explicitly states the inability to answer based on available information, acknowledging that the document does not mention the original author, and provides alternative context appropriate to the document's content (intangible cultural heritage).	No Discussion Needed	Siapa pengarang asli cerita Ande-Ande Lumut menurut dokumen?	Dokumen tidak menyebutkan siapa pengarang asli cerita Ande-Ande Lumut. Cerita ini merupakan bagian dari warisan budaya takbenda yang diwariskan secara lisan dari generasi ke generasi.
Q48	E10	E10	E9	Rejected	There is no substantive answer, references, or supporting chunks. The system provides a fallback without content, so evaluation sets an abstain.	Discussion Needed	Mengapa cerita rakyat seperti Ande-Ande Lumut masih relevan hingga kini?	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!
Q60	E7	E7	E8	Rejected	The question contains anachronism (modern technology element) irrelevant to the traditional story context; the model successfully corrects the assumption and returns to the narrative context.	Discussion Needed	Mengapa para Kleting menyeberang sungai tidak menyewa speedboat?	Para Kleting tidak menyeberang sungai menggunakan speedboat karena cerita Ande-Ande Lumut adalah dongeng tradisional yang berlatar waktu sebelum teknologi modern seperti speedboat ada. [...]

The full set of HITL evaluation results is attached in Appendix J to ensure transparency and enable further independent analysis.

7.8. Role of HITL in the Final Dataset

The HITL process provides a human-validated ground truth for evaluating faithful response generation in RAG systems, particularly within culturally grounded knowledge domains. By combining multi-evaluator judgments, a structured error taxonomy, and transparent documentation, the dataset establishes a robust foundation for future research on human-AI alignment and evaluation methodology.

8. Dataset Governance, Traceability, and Reusability

This section describes the dataset governance principles, traceability mechanisms, and reusability provisions applied in the construction and distribution of the HITL-based RAG evaluation dataset. These principles are designed to ensure that the dataset is not only reusable, but also auditable, verifiable, and interpretable by external researchers.

8.1. Dataset Governance

The dataset is constructed under the principle that every evaluation artifact must have a clear role, context, and relational meaning within the experimental pipeline. Accordingly, dataset governance is enforced through several core policies.

First, the separation of evaluation roles is explicitly maintained. The RAG system is responsible for generating candidate responses, the LLM Judge generates source-based references and chunk identifiers, automatic metrics function as a screening layer, and final decisions are made exclusively by human evaluators through the HITL process. No single component acts as the sole determinant of answer quality.

Second, a strict distinction between system-level failure and content-level error is enforced as a core governance policy. Technical failures are recorded as system-level failures and are not conflated with content quality analysis. This separation prevents evaluation bias and ensures that analyses of RAG system behavior focus on model performance rather than infrastructure-related disruptions.

Third, the dataset enforces a single primary decision per record, expressed through one decision outcome and one error code. This rule maintains annotation consistency and avoids unplanned multi-label interpretations.

8.2. Traceability and Auditability

Traceability is a central design principle of the dataset. Each data record is structured to allow explicit tracing across all components involved in the evaluation process.

Every dataset entry links the query, candidate response, reference response, and chunk identifier within the same record. As a result, every evaluation decision, whether produced by human judgment or supported by automatic metrics, can be traced back to a specific knowledge source.

The use of chunk identifiers as atomic reference units enables more precise auditing compared to document-level attribution. External researchers can directly inspect the exact source text underlying an evaluation decision without reinterpreting broader document context.

The n8n workflow serves as a shared execution context that preserves query consistency across all stages of the pipeline. This mechanism ensures that all evaluation artifacts are generated from the same query input and remain aligned throughout execution, reinforcing the integrity of inter-field relationships within the dataset.

8.3. Dataset Versioning and Reproducibility

The dataset is released under an explicit versioning scheme to support reproducibility and controlled reuse. Each dataset version reflects:

- the set of queries used,
- the version of cultural documents and derived chunks,
- the prompt configurations for the LLM RAG and LLM Judge models, and
- the HITL evaluation policy in effect.

Any change to these components results in a new dataset version, enabling transparent cross-version comparison. Version information and major changes are explicitly documented in the dataset metadata accompanying each public release.

Although the experimental pipeline involves dynamic components such as large language models, the dataset itself is provided as an evaluation snapshot that captures the concrete outputs of a specific experimental configuration. This approach allows comparative analysis without assuming full replication of runtime conditions.

8.4. Reusability and Scope of Use

The dataset is designed to support reuse across multiple research contexts, particularly for:

- evaluating faithfulness and grounding in RAG systems,
- comparative studies between automatic metrics and human evaluation, and
- analysis of hallucination behavior in LLM-based systems.

However, the dataset is not intended as a general-purpose performance benchmark across models or systems. All evaluation results are contextualized by:

- the cultural knowledge domain employed,
- the query design, and
- the applied HITL evaluation policy.

Users of the dataset are expected to explicitly reference this documentation when using the dataset for analysis or comparison and to avoid interpreting the results outside the experimental context described.

8.5. Ethical Considerations and Limitations

The dataset is constructed without the use of personal data or sensitive information. All cultural documents included are either publicly available or curated specifically for research purposes.

The primary limitation of the dataset lies in its dependence on model configurations and prompt designs at the time of experimentation. Consequently, evaluation outcomes cannot be directly generalized to different system configurations without additional analysis.

Nevertheless, by providing structured and traceable evaluation artifacts, the dataset is intended to serve as a methodological reference for future research on RAG evaluation and HITL-based assessment.

Closing Remarks

By explicitly applying principles of governance, traceability, and reusability, this dataset provides a transparent and accountable evaluation foundation for HITL-based RAG studies. The documentation complements the main journal article by offering operational details that cannot be fully accommodated within the primary publication.

9. Conclusion and Future Work

This work presents a structured and transparent HITL-based evaluation dataset designed to support rigorous assessment of faithfulness and grounding in Retrieval-Augmented Generation systems within a culturally grounded knowledge domain. By explicitly separating candidate generation, source-based reference construction, automatic screening, and human judgment, the dataset establishes a clear and auditable evaluation pipeline that avoids common ambiguities in RAG evaluation.

A central contribution of this dataset lies in its emphasis on traceability and role separation. Queries, candidate responses, reference outputs, chunk-level grounding, automatic metrics, and human decisions are aligned at the record level, enabling fine-grained inspection of model behavior and evaluation outcomes. The use of chunk identifiers as atomic grounding units further strengthens auditability and supports reproducible analysis beyond document-level attribution.

The Human-in-the-Loop framework serves as the final authority for evaluation, supported rather than replaced by automatic metrics. The multi-evaluator design, structured error taxonomy, and explicit handling of system-level versus content-level failures together provide a robust foundation for analyzing both correct behavior and failure modes of RAG systems. In this respect, the dataset functions not only as an evaluation resource, but also as a methodological reference for HITL-oriented assessment design.

Despite these strengths, the dataset reflects a snapshot of a specific experimental configuration. Evaluation outcomes are inherently dependent on the selected cultural domain, query design, chunking strategy, and model prompt configurations. As such, the dataset is not intended as a universal benchmark, but rather as a controlled and well-documented evaluation artifact that can be reused, extended, or adapted with clear awareness of its scope and limitations.

Future work may proceed along several directions. First, the dataset can be expanded to include additional cultural domains or multilingual settings, enabling comparative studies of grounding behavior across knowledge contexts. Second, alternative chunking strategies and memory injection schemes may be evaluated using the same HITL framework to assess their impact on faithfulness and hallucination patterns. Third, longitudinal versions of the dataset may be constructed to analyze how changes in model architectures or prompting strategies influence evaluation outcomes over time.

More broadly, this dataset highlights the continued importance of human judgment in the evaluation of grounded language generation. While automatic metrics provide useful diagnostic signals, reliable assessment of faithfulness, attribution, and contextual relevance remains fundamentally human-centered. By making these evaluation processes explicit and traceable, the dataset aims to support future research on human-AI alignment, evaluation methodology, and responsible deployment of RAG-based systems.

Appendix A: Overview of Curated Cultural Documents (Original Indonesian)

Table 2 - Overview of Curated Cultural Documents (Original Indonesian)

Penjelasan Tambahan

Bagian paragraf ini disusun sebagai penjelasan konseptual untuk menggambarkan potensi kesenjangan antara keselarasan semantik dan kepatuhan terhadap aturan evaluasi dalam sistem generatif berbasis konteks. Dalam beberapa kondisi, model dapat menghasilkan jawaban dengan tingkat similarity yang tinggi dan indikasi halusinasi yang rendah, sehingga secara permukaan tampak selaras dengan konteks yang diberikan. Namun, keselarasan tersebut tidak selalu menjamin bahwa jawaban dihasilkan dengan mematuhi aturan grounding dan penggunaan konteks secara eksplisit. Oleh karena itu, kondisi ini menjadi pertimbangan penting dalam evaluasi human-judge atau penilaian semantik oleh manusia untuk memastikan bahwa jawaban tidak hanya tampak benar secara semantik, tetapi juga sesuai dengan aturan dan tujuan evaluasi yang ditetapkan. Penjelasan ini digunakan untuk memperkuat rasionalitas perlunya keterlibatan manusia dalam proses evaluasi, terutama dalam membedakan keselarasan semu dan kepatuhan kontekstual yang sesungguhnya.

Ande-Ande Lumut: Mahakarya Takbenda dalam Mozaik Budaya Nusantara

Ande-Ande Lumut bukan sekadar dongeng pengantar tidur; ia adalah sebuah mahakarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa. Kisah ini telah melampaui batasan waktu dan ruang, membuktikan kekuatan narasi lisan dalam membentuk moral, etika, dan identitas sebuah bangsa. Dari pelosok desa hingga panggung-panggung modern, Ande-Ande Lumut terus dituturkan, diperankan, dan dihayati, menjadikannya salah satu pilar penting dalam warisan budaya takbenda Nusantara yang patut terus dilestarikan.

Kisah Klasik yang Memikat: Intisari Cerita Ande-Ande Lumut

Pada intinya, Ande-Ande Lumut adalah epos tentang pencarian jodoh seorang pangeran yang bijaksana dengan cara yang tak terduga. Alkisah, Pangeran Ande-Ande Lumut, seorang putra mahkota yang memegang teguh nilai-nilai kebaikan, mengumumkan sayembara untuk mencari calon permaisuri. Berbagai gadis cantik dari penjuru negeri berbondong-bondong datang, berharap dapat memikat hati sang pangeran. Mereka berdandan secantik mungkin, menunjukkan segala kelebihan fisik dan kemewahan yang mereka miliki. Namun, Ande-Ande Lumut menolak semua tawaran itu, seolah ia tengah mencari sesuatu yang lebih dari sekadar rupa.

Fokus pencariannya tertuju pada satu nama: Kleting Kuning. Kleting Kuning adalah antitesis dari para gadis pelamar lainnya. Ia digambarkan sebagai gadis yang miskin, lusuh, dan seringkali kotor karena pekerjaan kasar yang dibebankan oleh ibu tiri dan saudara-saudara tirinya: Kleting Merah, Kleting Biru, dan Kleting Hijau yang semua kejam. Kehidupan Kleting Kuning penuh dengan penderitaan dan penindasan, namun ia memiliki hati yang tulus, sabar, dan takwa.

Perjalanan para Kleting menuju sayembara Ande-Ande Lumut melibatkan sebuah rintangan krusial: penyeberangan sungai yang dijaga oleh sosok mitologis bernama Yuyu Kangkang. Yuyu Kangkang, seekor kepiting raksasa penjaga sungai, menawarkan jasa penyeberangan dengan imbalan ciuman dari setiap gadis. Para Kleting lainnya, yang terburu-buru dan mementingkan penampilan serta tujuan mereka, rela mencium Yuyu Kangkang demi bisa menyeberang, menunjukkan betapa rendahnya nilai moralitas mereka. Namun, Kleting Kuning, dengan keberanian dan harga diri yang tinggi, menolak mentah-mentah permintaan kotor itu. Ia justru dengan tegas menendang Yuyu Kangkang hingga tak berdaya, lalu menyeberangi sungai dengan kekuatannya sendiri, sebuah tindakan yang mencerminkan keteguhan prinsipnya.

Setibanya di istana Ande-Ande Lumut, satu per satu Kleting Merah, Biru, dan Hijau tampil di hadapan pangeran. Namun, Ande-Ande Lumut, yang memiliki kearifan luar biasa, menolak mereka semua, bahkan menyebutkan aroma amis Yuyu Kangkang yang melekat pada diri mereka. Barulah ketika Kleting Kuning tiba, dengan pakaian compang-camping dan tubuh yang mungkin masih kotor akibat perjalanan, Ande-Ande Lumut menyambutnya dengan sukacita. Pangeran telah mengetahui identitas asli Kleting Kuning dan bukanlah gadis miskin biasa, melainkan seorang putri yang menyamar atau berada dalam penyamaran takdir. Dengan penerimaan ini, cerita mencapai puncaknya, di mana kebaikan hati dan kemurnian jiwa Kleting Kuning diakui dan diganjar dengan kebahagiaan abadi bersama Ande-Ande Lumut.

Kedalaman Makna dan Pesan Moral: Cerminan Nilai Budaya

Di balik alur ceritanya yang sederhana, Ande-Ande Lumut mengandung lapisan makna dan pesan moral yang sangat dalam, menjadikannya instrumen efektif dalam pendidikan karakter dan pewarisan nilai-nilai luhur.

Pertama dan yang paling utama, cerita ini mengajarkan prioritas budi pekerti di atas penampilan fisik. Ande-Ande Lumut secara eksplisit menolak kecantikan lahiriah para Kleting lain yang hanya mementingkan rupa dan status, sementara ia memilih Kleting Kuning yang lusuh namun berhati mulia. Ini adalah kritik tajam terhadap superficialitas dan penekanan pada esensi sejati manusia dengan karakter, kejujuran, dan kebaikan hati. Pesan ini relevan sepanjang masa, mengingatkan bahwa nilai seseorang tidak ditentukan oleh kemewahan atau kecantikan fisik semata, melainkan oleh integritas dan akhlaknya.

Kedua, kisah ini mengilustrasikan pentingnya kesabaran dan ketabahan dalam menghadapi cobaan. Kleting Kuning adalah epitome dari figur yang sabar. Ia menghadapi perlakuan kejam dari ibu dan saudara tirinya tanpa pernah membalas dendam atau menyerah pada nasib. Kesabarannya dalam menanggung penderitaan dan ketabahannya dalam menghadapi rintangan, seperti Yuyu Kangkang, akhirnya membuahkan hasil yang manis. Ini adalah pengingat bahwa ketekunan dan ketabahan akan membawa seseorang pada kebahagiaan dan keadilan.

Ketiga, Ande-Ande Lumut juga menyoroti keberanian dan integritas. Penolakan Kleting Kuning untuk mencium Yuyu Kangkang adalah simbol kuat dari penolakannya terhadap kompromi moral. Ia tidak mau menukar kehormatan dan harga dirinya demi kemudahan sesaat. Tindakan ini membedakannya dari saudara-saudara tirinya yang rela mengorbankan martabat

demi mencapai tujuan. Pesan ini menekankan pentingnya memegang teguh prinsip kebenaran dan kejujuran, bahkan ketika dihadapkan pada godaan atau tekanan.

Ande-Ande Lumut sebagai Warisan Budaya Takbenda: Sebuah Jembatan Antar Generasi

Ande-Ande Lumut bukan sekadar cerita yang tercetak di lembaran buku; ia adalah warisan budaya takbenda (intangible cultural heritage) yang hidup dan terus berevolusi. Statusnya sebagai warisan takbenda berarti bahwa nilai dan keberadaannya tidak terletak pada objek fisik, melainkan pada praktik, pengetahuan, tradisi lisan, ekspresi, dan keterampilan yang diwariskan dari satu generasi ke generasi berikutnya.

Pada masa lampau, jauh sebelum era digital dan media massa, Ande-Ande Lumut adalah bagian integral dari kehidupan sehari-hari masyarakat. Kisah ini dituturkan secara lisan di balai-balai desa, di teras rumah saat senja, atau sebagai pengantar tidur bagi anak-anak. Para penutur, seringkali sesepuh atau juru dongeng, tidak hanya sekadar menyampaikan plot, melainkan juga menghidupkan cerita dengan intonasi yang dramatis, ekspresi wajah yang beragam, dan gerak tubuh yang ekspresif. Setiap penuturan adalah sebuah pertunjukan mini yang memungkinkan audiens untuk terhanyut dalam imajinasi dan menyerap nilai-nilai yang disampaikan.

Lebih dari itu, Ande-Ande Lumut juga menemukan panggungnya dalam seni pertunjukan tradisional. Di Jawa, kisah ini sering diadaptasi dan dipentaskan dalam ketoprak dan wayang orang. Dalam pertunjukan-pertunjukan ini, karakter-karakter Ande-Ande Lumut divisualisasikan melalui kostum yang indah, tata rias yang khas, dan gerakan tari yang anggun. Dialog-dialognya diiringi oleh alunan gamelan yang magis, menciptakan atmosfer yang memukau dan pengalaman budaya yang mendalam. Melalui medium ini, cerita tidak hanya disimak, tetapi juga disaksikan dan dirasakan secara kolektif, memperkuat ikatan komunitas dan identitas budaya.

Di era modern saat ini, Ande-Ande Lumut telah beradaptasi dengan perkembangan teknologi tanpa kehilangan esensinya. Kisah ini kini banyak ditemukan dalam buku cerita anak-anak bergambar, yang memungkinkan anak-anak untuk secara visual menjelajahi dunia Ande-Ande Lumut. Platform digital seperti YouTube dan aplikasi dongeng juga menjadi medium baru bagi cerita ini dalam bentuk animasi, film pendek, atau audiodrama, menjangkau audiens global dan generasi yang lebih akrab dengan teknologi. Adaptasi-adaptasi ini membantu menjaga relevansi cerita di tengah derasnya arus informasi dan hiburan modern.

Meskipun demikian, bentuk-bentuk penuturan tradisional tetap lestari. Banyak sanggar seni, sekolah, dan komunitas budaya masih aktif mementaskan Ande-Ande Lumut dalam bentuk ketoprak atau wayang orang, seringkali dengan sentuhan kontemporer untuk menarik minat audiens muda. Bahkan, Ande-Ande Lumut juga seringkali menjadi bagian dari kurikulum pendidikan, memastikan bahwa kekayaan narasi ini terus dikenal dan dipelajari oleh generasi penerus, tidak hanya sebagai hiburan tetapi sebagai materi pembelajaran yang kaya akan nilai dan kearifan lokal.

Sebagai penutup, Ande-Ande Lumut adalah bukti nyata bagaimana sebuah cerita rakyat mampu menjadi jembatan abadi yang menghubungkan masa lalu, kini, dan masa depan. Ia

adalah penjaga nilai-nilai luhur, pengingat akan pentingnya budi pekerti, kesabaran, dan integritas. Melalui keberlanjutan penuturannya dalam berbagai bentuk mulai dari lisan tradisional hingga adaptasi digital ceritera Ande-Ande Lumut terus berdenyut, menginspirasi, dan memperkaya mozaik budaya Nusantara yang begitu beragam. Melestarikannya berarti melestarikan identitas dan kearifan yang tak ternilai bagi generasi mendatang.

Appendix B: Chunk Schema

Table 3 – The Completed Chunk Schema

Document_ID	Index	Chunk_ID	Chunk_Text (Original Indonesian)
Doc_01	1	chunk-0	Bagian paragraf ini disusun sebagai penjelasan konseptual untuk menggambarkan potensi kesenjangan antara keselarasan semantik dan kepatuhan terhadap aturan evaluasi dalam sistem generatif berbasis konteks. Dalam beberapa kondisi, model dapat menghasilkan jawaban dengan tingkat similarity yang tinggi dan indikasi halusinasi yang rendah, sehingga secara permukaan tampak selaras dengan konteks yang diberikan. Namun, keselarasan tersebut tidak selalu menjamin bahwa jawaban dihasilkan dengan mematuhi aturan grounding dan penggunaan konteks secara eksplisit. Oleh karena itu, kondisi ini menjadi pertimbangan penting dalam evaluasi human-judge atau penilaian semantik oleh manusia untuk memastikan bahwa jawaban tidak hanya tampak benar secara semantik, tetapi juga ses
	2	chunk-500	nting dalam evaluasi human-judge atau penilaian semantik oleh manusia untuk memastikan bahwa jawaban tidak hanya tampak benar secara semantik, tetapi juga sesuai dengan aturan dan tujuan evaluasi yang ditetapkan. Penjelasan ini digunakan untuk memperkuat rasionalitas perlunya keterlibatan manusia dalam proses evaluasi, terutama dalam membedakan keselarasan semu dan kepatuhan kontekstual yang sesungguhnya. Ande-Ande Lumut: Mahakarya Takbenda dalam Mozaik Budaya Nusantara Ande-Ande Lumut bukan sekadar dongeng pengantar tidur; ia adalah sebuah mahakarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa. Kisah ini telah melampaui batasan waktu dan ruang, membuktikan kekuatan narasi lisan dalam membentuk moral, etika, dan i
	3	chunk-1000	akarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa. Kisah ini telah melampaui batasan waktu dan ruang, membuktikan kekuatan narasi lisan dalam membentuk moral, etika, dan identitas sebuah bangsa. Dari pelosok desa hingga panggung-panggung modern, Ande-Ande Lumut terus dituturkan, diperankan, dan dihayati, menjadikannya salah satu pilar penting dalam warisan budaya takbenda Nusantara yang patut terus dilestarikan. Kisah Klasik yang Memikat: Intisari Cerita Ande-Ande Lumut Pada intinya, Ande-Ande Lumut adalah epos tentang pencarian jodoh seorang pangeran yang bijaksana dengan cara yang tak terduga. Alkisah, Pangeran Ande-Ande Lumut, seorang putra mahkota yang memega
	4	chunk-1500	mut Pada intinya, Ande-Ande Lumut adalah epos tentang pencarian jodoh seorang pangeran yang bijaksana dengan

Document_ID	Index	Chunk_ID	Chunk_Text (Original Indonesian)
			cara yang tak terduga. Alkisah, Pangeran Ande-Ande Lumut, seorang putra mahkota yang memegang teguh nilai-nilai kebaikan, mengumumkan sayembara untuk mencari calon permaisuri. Berbagai gadis cantik dari penjuru negeri berbondong-bondong datang, berharap dapat memikat hati sang pangeran. Mereka berdandan secantik mungkin, menunjukkan segala kelebihan fisik dan kemewahan yang mereka miliki. Namun, Ande-Ande Lumut menolak semua tawaran itu, seolah ia tengah mencari sesuatu yang lebih dari sekadar rupa. Fokus pencariannya tertuju pada satu nama: Kleting Kuning. Kleting Kuning adalah antit
	5	chunk-2000	Namun, Ande-Ande Lumut menolak semua tawaran itu, seolah ia tengah mencari sesuatu yang lebih dari sekadar rupa. Fokus pencariannya tertuju pada satu nama: Kleting Kuning. Kleting Kuning adalah antitesis dari para gadis pelamar lainnya. Ia digambarkan sebagai gadis yang miskin, lusuh, dan seringkali kotor karena pekerjaan kasar yang dibebankan oleh ibu tiri dan saudara-saudara tirinya— Kleting Merah, Kleting Biru, dan Kleting Hijau— yang kejam. Kehidupan Kleting Kuning penuh dengan penderitaan dan penindasan, namun ia memiliki hati yang tulus, sabar, dan takwa. Perjalanan para Kleting menuju sayembara Ande-Ande Lumut melibatkan sebuah rintangan krusial: penyeberangan sungai yang dijaga oleh
	6	chunk-2500	an penindasan, namun ia memiliki hati yang tulus, sabar, dan takwa. Perjalanan para Kleting menuju sayembara Ande-Ande Lumut melibatkan sebuah rintangan krusial: penyeberangan sungai yang dijaga oleh sosok mitologis bernama Yuyu Kangkang. Yuyu Kangkang, seekor kepiting raksasa penjaga sungai, menawarkan jasa penyeberangan dengan imbalan ciuman dari setiap gadis. Para Kleting lainnya, yang terburu-buru dan mementingkan penampilan serta tujuan mereka, rela mencium Yuyu Kangkang demi bisa menyeberang, menunjukkan betapa rendahnya nilai moralitas mereka. Namun, Kleting Kuning, dengan keberanian dan harga diri yang tinggi, menolak mentah-mentah permintaan kotor itu. Ia justru dengan tegas menend
	7	chunk-3000	ng, menunjukkan betapa rendahnya nilai moralitas mereka. Namun, Kleting Kuning, dengan keberanian dan harga diri yang tinggi, menolak mentah-mentah permintaan kotor itu. Ia justru dengan tegas menendang Yuyu Kangkang hingga tak berdaya, lalu menyeberangi sungai dengan kekuatannya sendiri, sebuah tindakan yang mencerminkan keteguhan prinsipnya. Setibanya di istana Ande-Ande Lumut, satu per satu Kleting Merah, Biru, dan Hijau tampil di hadapan pangeran. Namun, Ande-Ande Lumut, yang memiliki kearifan luar biasa, menolak mereka semua, bahkan menyebutkan aroma amis Yuyu Kangkang

Document_ID	Index	Chunk_ID	Chunk_Text (Original Indonesian)
			yang melekat pada diri mereka. Barulah ketika Kleting Kuning tiba, dengan pakaian compang-camping dan tubuh yang mung
	8	chunk-3500	fan luar biasa, menolak mereka semua, bahkan menyebutkan aroma amis Yuyu Kangkang yang melekat pada diri mereka. Barulah ketika Kleting Kuning tiba, dengan pakaian compang-camping dan tubuh yang mungkin masih kotor akibat perjalanan, Ande-Ande Lumut menyambutnya dengan sukacita. Pangeran telah mengetahui identitas asli Kleting Kuning—bukanlah gadis miskin biasa, melainkan seorang putri yang menyamar atau berada dalam penyamaran takdir. Dengan penerimaan ini, cerita mencapai puncaknya, di mana kebaikan hati dan kemurnian jiwa Kleting Kuning diakui dan diganjar dengan kebahagiaan abadi bersama Ande-Ande Lumut. Kedalaman Makna dan Pesan Moral: Cerminan Nilai Budaya Di balik alur ceritanya yang
	9	chunk-4000	baikan hati dan kemurnian jiwa Kleting Kuning diakui dan diganjar dengan kebahagiaan abadi bersama Ande-Ande Lumut. Kedalaman Makna dan Pesan Moral: Cerminan Nilai Budaya Di balik alur ceritanya yang sederhana, Ande-Ande Lumut mengandung lapisan makna dan pesan moral yang sangat dalam, menjadikannya instrumen efektif dalam pendidikan karakter dan pewarisan nilai-nilai luhur. Pertama dan yang paling utama, cerita ini mengajarkan prioritas budi pekerti di atas penampilan fisik. Ande-Ande Lumut secara eksplisit menolak kecantikan lahiriah para Kleting lain yang hanya mementingkan rupa dan status, sementara ia memilih Kleting Kuning yang lusuh namun berhati mulia. Ini adalah kritik tajam terhadap
	10	chunk-4500	ara eksplisit menolak kecantikan lahiriah para Kleting lain yang hanya mementingkan rupa dan status, sementara ia memilih Kleting Kuning yang lusuh namun berhati mulia. Ini adalah kritik tajam terhadap superficialitas dan penekanan pada esensi sejati manusia—yaitu karakter, kejujuran, dan kebaikan hati. Pesan ini relevan sepanjang masa, mengingatkan bahwa nilai seseorang tidak ditentukan oleh kemewahan atau kecantikan fisik semata, melainkan oleh integritas dan akhlakunya. Kedua, kisah ini mengilustrasikan pentingnya kesabaran dan ketabahan dalam menghadapi cobaan. Kleting Kuning adalah epitome dari figur yang sabar. Ia menghadapi perlakuan kejam dari ibu dan saudara tirinya tanpa pernah memb
	11	chunk-5000	ustrasikan pentingnya kesabaran dan ketabahan dalam menghadapi cobaan. Kleting Kuning adalah epitome dari figur yang sabar. Ia menghadapi perlakuan kejam dari ibu dan saudara tirinya tanpa pernah membalas dendam atau menyerah pada nasib. Kesabarannya dalam menanggung penderitaan dan ketabahannya dalam menghadapi rintangan, seperti Yuyu Kangkang, akhirnya membuahkan

Document_ID	Index	Chunk_ID	Chunk_Text (Original Indonesian)
			hasil yang manis. Ini adalah pengingat bahwa ketekunan dan ketabahan akan membawa seseorang pada kebahagiaan dan keadilan. Ketiga, Ande-Ande Lumut juga menyoroti keberanian dan integritas. Penolakan Kleting Kuning untuk mencium Yuyu Kangkang adalah simbol kuat dari penolakannya terhadap kompromi moral. Ia tidak mau menukar kehorma
	12	chunk-5500	nde-Ande Lumut juga menyoroti keberanian dan integritas. Penolakan Kleting Kuning untuk mencium Yuyu Kangkang adalah simbol kuat dari penolakannya terhadap kompromi moral. Ia tidak mau menukar kehormatan dan harga dirinya demi kemudahan sesaat. Tindakan ini membedakannya dari saudara-saudara tirinya yang rela mengorbankan martabat demi mencapai tujuan. Pesan ini menekankan pentingnya memegang teguh prinsip kebenaran dan kejujuran, bahkan ketika dihadapkan pada godaan atau tekanan. Ande-Ande Lumut sebagai Warisan Budaya Takbenda: Sebuah Jembatan Antar Generasi Ande-Ande Lumut bukan sekadar cerita yang tercetak di lembaran buku; ia adalah warisan budaya takbenda (intangible cultural heritage)
	13	chunk-6000	t sebagai Warisan Budaya Takbenda: Sebuah Jembatan Antar Generasi Ande-Ande Lumut bukan sekadar cerita yang tercetak di lembaran buku; ia adalah warisan budaya takbenda (intangible cultural heritage) yang hidup dan terus berevolusi. Statusnya sebagai warisan takbenda berarti bahwa nilai dan keberadaannya tidak terletak pada objek fisik, melainkan pada praktik, pengetahuan, tradisi lisan, ekspresi, dan keterampilan yang diwariskan dari satu generasi ke generasi berikutnya. Pada masa lampau, jauh sebelum era digital dan media massa, Ande-Ande Lumut adalah bagian integral dari kehidupan sehari-hari masyarakat. Kisah ini dituturkan secara lisan di balai-balai desa, di teras rumah saat senja, at
	14	chunk-6500	sebelum era digital dan media massa, Ande-Ande Lumut adalah bagian integral dari kehidupan sehari-hari masyarakat. Kisah ini dituturkan secara lisan di balai-balai desa, di teras rumah saat senja, atau sebagai pengantar tidur bagi anak-anak. Para penutur, seringkali sesepuh atau juru dongeng, tidak hanya sekadar menyampaikan plot, melainkan juga menghidupkan cerita dengan intonasi yang dramatis, ekspresi wajah yang beragam, dan gerak tubuh yang ekspresif. Setiap penuturan adalah sebuah pertunjukan mini yang memungkinkan audiens untuk terhanyut dalam imajinasi dan menyerap nilai-nilai yang disampaikan. Lebih dari itu, Ande-Ande Lumut juga menemukan panggungnya dalam seni pertunjukan tradisio
	15	chunk-7000	kan mini yang memungkinkan audiens untuk terhanyut dalam imajinasi dan menyerap nilai-nilai yang

Document_ID	Index	Chunk_ID	Chunk_Text (Original Indonesian)
			disampaikan. Lebih dari itu, Ande-Ande Lumut juga menemukan panggungnya dalam seni pertunjukan tradisional. Di Jawa, kisah ini sering diadaptasi dan dipentaskan dalam ketoprak dan wayang orang. Dalam pertunjukan-pertunjukan ini, karakter-karakter Ande-Ande Lumut divisualisasikan melalui kostum yang indah, tata rias yang khas, dan gerakan tari yang anggun. Dialog-dialognya diiringi oleh alunan gamelan yang magis, menciptakan atmosfer yang memukau dan pengalaman budaya yang mendalam. Melalui medium ini, cerita tidak hanya disimak, tetapi juga disaksikan dan dirasakan secara kolektif, memperkuat ika
	16	chunk-7500	yang magis, menciptakan atmosfer yang memukau dan pengalaman budaya yang mendalam. Melalui medium ini, cerita tidak hanya disimak, tetapi juga disaksikan dan dirasakan secara kolektif, memperkuat ikatan komunitas dan identitas budaya. Di era modern saat ini, Ande-Ande Lumut telah beradaptasi dengan perkembangan teknologi tanpa kehilangan esensinya. Kisah ini kini banyak ditemukan dalam buku cerita anak-anak bergambar, yang memungkinkan anak-anak untuk secara visual menjelajahi dunia Ande-Ande Lumut. Platform digital seperti YouTube dan aplikasi dongeng juga menjadi medium baru bagi cerita ini dalam bentuk animasi, film pendek, atau audiodrama, menjangkau audiens global dan generasi yang leb
	17	chunk-8000	umut. Platform digital seperti YouTube dan aplikasi dongeng juga menjadi medium baru bagi cerita ini dalam bentuk animasi, film pendek, atau audiodrama, menjangkau audiens global dan generasi yang lebih akrab dengan teknologi. Adaptasi-adaptasi ini membantu menjaga relevansi cerita di tengah derasnya arus informasi dan hiburan modern. Meskipun demikian, bentuk-bentuk penuturan tradisional tetap lestari. Banyak sanggar seni, sekolah, dan komunitas budaya masih aktif mementaskan Ande-Ande Lumut dalam bentuk ketoprak atau wayang orang, seringkali dengan sentuhan kontemporer untuk menarik minat audiens muda. Bahkan, Ande-Ande Lumut juga seringkali menjadi bagian dari kurikulum pendidikan, memast
	18	chunk-8500	lam bentuk ketoprak atau wayang orang, seringkali dengan sentuhan kontemporer untuk menarik minat audiens muda. Bahkan, Ande-Ande Lumut juga seringkali menjadi bagian dari kurikulum pendidikan, memastikan bahwa kekayaan narasi ini terus dikenal dan dipelajari oleh generasi penerus, tidak hanya sebagai hiburan tetapi sebagai materi pembelajaran yang kaya akan nilai dan kearifan lokal. Sebagai penutup, Ande-Ande Lumut adalah bukti nyata bagaimana sebuah cerita rakyat mampu menjadi jembatan abadi yang menghubungkan masa lalu,

Document_ID	Index	Chunk_ID	Chunk_Text (Original Indonesian)
			kini, dan masa depan. Ia adalah penjaga nilai-nilai luhur, pengingat akan pentingnya budi pekerti, kesabaran, dan integritas. Melalui keberlanjutan penuturannya dalam berb
	19	chunk-9000	ang menghubungkan masa lalu, kini, dan masa depan. Ia adalah penjaga nilai-nilai luhur, pengingat akan pentingnya budi pekerti, kesabaran, dan integritas. Melalui keberlanjutan penuturannya dalam berbagai bentuk—dari lisan tradisional hingga adaptasi digital—Ande-Ande Lumut terus berdenyut, menginspirasi, dan memperkaya mozaik budaya Nusantara yang begitu beragam. Melestarikannya berarti melestarikan identitas dan kearifan yang tak ternilai bagi generasi mendatang.
	20	chunk-9500	Additional chunk. Untuk mengarahkan response jika tidak menemukan teks relevan dari chunk 0-9000. Seperti Maaf tidak menemukan chunks yang relevan untuk menjawab pertanyaan. System error trim.text. Nyatakan dengan Chunk-9500.

Clarification on Ground Truth Scope (Appendix B)

Within the chunk schema presented in Appendix B, not all chunks function as domain-level ground truth. Specifically, *chunk-0* and *chunk-500* contain methodological and conceptual explanations that are intentionally included to support evaluation rationale and Human-in-the-Loop (HITL) decision alignment. These chunks are **not used as factual grounding references** for answering user queries.

Furthermore, *chunk-9500* is explicitly designed as a **failure-trapping and system fallback chunk**, activated only when no relevant knowledge chunks are retrieved or when the system must generate an abstention. This chunk serves a **control and traceability function** within the evaluation pipeline and does **not constitute ground truth content**. All content-level faithfulness assessments are therefore conducted **exclusively** with respect to chunks representing curated cultural knowledge.

Appendix C: The Completed Query Categories and Texts

Table 4 – The Completed Query Categories and Texts

Query_ID	Category	Query_Text (Original Indonesian)
Q1	In-context	Mengapa Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau?
Q2	In-context	Bagaimana sikap Kleting Kuning sepanjang perjalanan?
Q3	In-context	Apa peran Yuyu Kangkang dalam alur cerita?
Q4	In-context	Bagaimana ujian sungai digambarkan dalam cerita?
Q5	In-context	Apa yang membedakan Kleting Kuning dari saudara-saudaranya?
Q6	Ambiguous	Apa makna kisah Ande-Ande Lumut bagi kehidupan manusia?
Q7	Ambiguous	Nilai apa yang paling menonjol dalam cerita ini?
Q8	Ambiguous	Bagaimana kisah ini dapat dimaknai secara budaya?
Q9	Ambiguous	Apa pesan utama perjalanan Kleting Kuning?
Q10	Ambiguous	Mengapa cerita Ande-Ande Lumut tetap relevan hingga kini?
Q11	Out-of-context	Siapa pengarang asli cerita Ande-Ande Lumut menurut dokumen?
Q12	Out-of-context	Pada tahun berapa cerita ini pertama kali ditulis?
Q13	Out-of-context	Apakah cerita ini berasal dari Kerajaan Majapahit?
Q14	Out-of-context	Berapa jumlah pementasan Ande-Ande Lumut di Jawa?
Q15	Out-of-context	Apakah kisah ini pernah difilmkan menurut dokumen?
Q16	In-context	Mengapa Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau?
Q17	In-context	Bagaimana sikap Kleting Kuning sepanjang perjalanan?
Q18	In-context	Apa peran Yuyu Kangkang dalam alur cerita?
Q19	In-context	Bagaimana ujian sungai digambarkan dalam cerita?
Q20	In-context	Apa yang membedakan Kleting Kuning dari saudara-saudaranya?
Q21	Ambiguous	Apa makna kisah Ande-Ande Lumut bagi kehidupan manusia?
Q22	Ambiguous	Nilai apa yang paling menonjol dalam cerita ini?
Q23	Ambiguous	Bagaimana kisah ini dapat dimaknai secara budaya?
Q24	Ambiguous	Apa pesan utama perjalanan Kleting Kuning?
Q25	Ambiguous	Mengapa cerita Ande-Ande Lumut tetap relevan hingga kini?
Q26	Out-of-context	Siapa pengarang asli cerita Ande-Ande Lumut menurut dokumen?
Q27	Out-of-context	Pada tahun berapa cerita ini pertama kali ditulis?
Q28	Out-of-context	Apakah cerita ini berasal dari Kerajaan Majapahit?
Q29	Out-of-context	Berapa jumlah pementasan Ande-Ande Lumut di Jawa?
Q30	In-context	Bagaimana kisah Ande-Ande Lumut berkontribusi dalam pembentukan nilai moral masyarakat Jawa?
Q31	In-context	Apa makna simbolik sosok Yuyu Kangkang dalam perjalanan para Kleting menuju sayembara Ande-Ande Lumut?
Q32	In-context	Mengapa kesabaran dan ketabahan Kleting Kuning dianggap sebagai teladan dalam cerita ini?
Q33	In-context	Bagaimana adaptasi modern cerita Ande-Ande Lumut membantu menjaga relevansinya di era digital?
Q34	In-context	Mengapa Ande-Ande Lumut disebut sebagai jembatan antar generasi dalam konteks warisan budaya takbenda?

Query_ID	Category	Query_Text (Original Indonesian)
Q35	In-context	Mengapa Ande-Ande Lumut memilih Kleting Kuning dibandingkan para Kleting lainnya?
Q36	In-context	Nilai moral apa yang terkandung dalam penolakan Kleting Kuning terhadap permintaan Yuyu Kangkang?
Q37	In-context	Bagaimana karakter kesabaran Kleting Kuning digambarkan dalam cerita Ande-Ande Lumut?
Q38	In-context	Bagaimana kisah Ande-Ande Lumut diwariskan dalam masyarakat Jawa menurut teks?
Q39	In-context	Mengapa Ande-Ande Lumut dianggap penting dalam konteks budaya Nusantara?
Q40	In-context	Bagaimana kisah Ande-Ande Lumut dapat digunakan sebagai media pembelajaran karakter di sekolah modern?
Q41	In-context	Apa relevansi cerita Ande-Ande Lumut terhadap tantangan moral generasi muda di era globalisasi?
Q42	In-context	Bagaimana peran cerita rakyat seperti Ande-Ande Lumut dalam memperkuat ketahanan budaya nasional?
Q43	In-context	Mengapa nilai kesabaran yang ditunjukkan Kleting Kuning penting dalam kehidupan sosial masyarakat saat ini?
Q44	In-context	Bagaimana adaptasi digital cerita Ande-Ande Lumut memengaruhi cara masyarakat memaknai budaya tradisional?
Q45	Ambiguous	Apa makna penting dari kisah Ande-Ande Lumut bagi masyarakat?
Q46	Ambiguous	Bagaimana seharusnya kisah Ande-Ande Lumut dipahami oleh pembaca?
Q47	Ambiguous	Nilai apa saja yang bisa diambil dari tokoh Kleting Kuning?
Q48	Ambiguous	Mengapa cerita rakyat seperti Ande-Ande Lumut masih relevan hingga kini?
Q49	Ambiguous	Apa pesan utama yang ingin disampaikan melalui cerita Ande-Ande Lumut?
Q50	In-context	Pada masa kerajaan apa kisah Ande-Ande Lumut pertama kali berkembang di Jawa?
Q51	In-context	Status sosial Ande-Ande Lumut sebagai pangeran berasal dari kerajaan mana menurut sejarah?
Q52	In-context	Apakah kisah Ande-Ande Lumut memiliki latar waktu sejarah tertentu yang dapat diidentifikasi secara kronologis?
Q53	In-context	Apakah tokoh Yuyu Kangkang terinspirasi dari makhluk mitologis tertentu dalam kepercayaan Jawa kuno?
Q54	In-context	Apakah cerita Ande-Ande Lumut pernah digunakan secara resmi sebagai alat propaganda budaya oleh pemerintah?
Q55	Out-of-context	Siapa tokoh sejarawan Jawa yang pertama kali mendokumentasikan kisah Ande-Ande Lumut dalam manuskrip tertulis?
Q56	Out-of-context	Sebutkan tahun dan lokasi ditemukannya naskah kuno Ande-Ande Lumut yang menjadi rujukan utama cerita ini.
Q57	Out-of-context	Bagaimana hubungan genealogis Ande-Ande Lumut dengan tokoh-tokoh kerajaan Jawa lainnya menurut catatan sejarah?
Q58	Out-of-context	Ritual budaya apa yang secara khusus terinspirasi dari kisah Yuyu Kangkang dalam tradisi Jawa kuno?

Query_ID	Category	Query_Text (Original Indonesian)
Q59	Out-of-context	Apakah dokumen ini menyebutkan versi terpendek atau terpanjang dari cerita Ande-Ande Lumut?
Q60	Out-of-context	Mengapa para Kleting menyeberang sungai tidak menyewa speedboad?
Q61	Ambiguous	Siapa yang menjadi pelanggan rutin menyeberang dan apakah mendapatkan potongan tarif?

Note on Repeated Queries.

Repeated queries are intentionally included to observe consistency across evaluation cycles. While identical queries may yield variations in candidate and reference responses due to the generative nature of LLMs, Human-in-the-Loop (HITL) judgments prioritize rule compliance and grounding fidelity, resulting in stable classification outcomes across repetitions.

Appendix D: Candidate and Reference Outputs (Original Indonesian)

Table 5 - Candidate and Reference Outputs (Original Indonesian)

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
Q1	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena mereka memiliki aroma amis Yuyu Kangkang yang melekat pada diri mereka. Ini menunjukkan bahwa mereka telah mengorbankan martabat demi kemudahan menyeberangi sungai, berbeda dengan Kleting Kuning yang berintegritas.	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena aroma amis Yuyu Kangkang yang melekat pada diri mereka.	chunk-3000
Q2	Sepanjang perjalanan, Kleting Kuning menunjukkan sikap tulus, sabar, dan takwa. Ia memiliki keberanian dan integritas tinggi, menolak mentah-mentah permintaan kotor Yuyu Kangkang dengan menendangnya, serta menyeberangi sungai dengan kekuatannya sendiri. Ia juga tabah menghadapi penderitaan.	Kleting Kuning memiliki hati yang tulus, sabar, dan takwa. Ia menunjukkan keberanian dan harga diri tinggi dengan menolak permintaan kotor Yuyu Kangkang, serta menendangnya hingga tak berdaya, mencerminkan keteguhan prinsipnya.	chunk-2000, chunk-2500, chunk-5000
Q3	Yuyu Kangkang berperan sebagai rintangan krusial dalam perjalanan para Kleting menuju sayembara Ande-Ande Lumut. Ia menguji moralitas para gadis dengan menawarkan penyeberangan sungai dengan imbalan ciuman. Penolakan Kleting Kuning terhadap Yuyu Kangkang menunjukkan keberanian dan integritasnya, yang menjadi pembeda utama di mata Pangeran Ande-Ande Lumut.	Yuyu Kangkang adalah kepiting raksasa penjaga sungai, menawarkan jasa penyeberangan dengan imbalan ciuman, menjadi rintangan krusial.	chunk-2500, chunk-3000
Q4	Ujian sungai dalam cerita Ande-Ande Lumut digambarkan sebagai rintangan krusial yang dijaga oleh Yuyu Kangkang, seekor kepiting raksasa. Ia menawarkan penyeberangan dengan imbalan ciuman. Gadis-gadis lain rela menciumnya, namun Kleting Kuning menolak tegas, menendang Yuyu Kangkang, dan menyeberang sendiri, menunjukkan keberanian serta integritasnya.	Ujian sungai melibatkan penyeberangan yang dijaga Yuyu Kangkang, seekor kepiting raksasa. Yuyu Kangkang menawarkan jasa penyeberangan dengan imbalan ciuman.	chunk-2500
Q5	Kleting Kuning berbeda dari saudara-saudaranya karena hati yang tulus, sabar, takwa, dan integritasnya. Ia menolak	Kleting Kuning memiliki hati yang tulus, sabar, dan takwa, menolak mencium Yuyu Kangkang, berbeda dengan	chunk-2000, chunk-2500,

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
	mengorbankan harga dirinya dengan mencium Yuyu Kangkang, tidak seperti saudara-saudaranya yang rela melakukannya. Ande-Ande Lumut memilihnya berdasarkan kemurnian jiwa, bukan penampilan fisik.	saudara-saudaranya yang mementingkan penampilan dan rela mencium Yuyu Kangkang.	chunk-3000, chunk-4000, chunk-4500, chunk-5000, chunk-5500
Q6	Kisah Ande-Ande Lumut mengajarkan bahwa budi pekerti, kesabaran, dan integritas jauh lebih berharga daripada penampilan fisik atau kekayaan. Cerita ini menekankan pentingnya karakter, kejujuran, dan kebaikan hati dalam menghadapi cobaan hidup, serta keberanian untuk memegang teguh prinsip moral.	Kisah Ande-Ande Lumut mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran, ketabahan, keberanian, dan integritas dalam menghadapi cobaan, serta memegang teguh prinsip kebenaran dan kejujuran.	chunk-4000, chunk-4500, chunk-5000, chunk-5500
Q7	Nilai paling menonjol dalam cerita Ande-Ande Lumut adalah prioritas budi pekerti di atas penampilan fisik, serta pentingnya kesabaran, ketabahan, keberanian, dan integritas dalam menghadapi cobaan hidup.	Prioritas budi pekerti di atas penampilan fisik adalah nilai paling menonjol dalam cerita ini.	chunk-4000, chunk-4500
Q8	Kisah Ande-Ande Lumut kaya akan makna budaya, terutama dalam menanamkan nilai-nilai luhur. Cerita ini menekankan pentingnya budi pekerti di atas penampilan fisik, mengajarkan bahwa kebaikan hati dan kejujuran lebih berharga daripada kecantikan lahiriah semata. Selain itu, kisah ini mengilustrasikan nilai kesabaran dan ketabahan dalam menghadapi kesulitan, serta keberanian untuk mempertahankan integritas dan harga diri. Makna-nilai ini menjadikan Ande-Ande Lumut sebagai warisan budaya takbenda yang berfungsi sebagai jembatan antar generasi, menjaga kearifan lokal dan identitas bangsa.	Di balik alur ceritanya yang sederhana, Ande-Ande Lumut mengandung lapisan makna dan pesan moral yang sangat dalam, menjadikannya instrumen efektif dalam pendidikan karakter dan pewarisan nilai-nilai luhur.	chunk-4000
Q9	Pesan utama dari perjalanan Kleting Kuning adalah penekanan pada nilai-nilai luhur seperti budi pekerti, kesabaran, ketabahan, keberanian, dan integritas di atas penampilan fisik atau kekayaan. Cerita ini mengajarkan bahwa kebaikan hati dan kemurnian jiwa akan selalu dihargai dan pada akhirnya akan	Perjalanan Kleting Kuning mengajarkan pentingnya kesabaran, ketabahan dalam menghadapi cobaan, keberanian, dan integritas.	chunk-5000, chunk-5500

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
	membawa kebahagiaan sejati, seperti yang ditunjukkan oleh Pangeran Ande-Ande Lumut yang memilih Kleting Kuning meskipun penampilannya lusuh.		
Q10	Cerita Ande-Ande Lumut tetap relevan karena mengandung pesan moral yang mendalam dan universal, seperti pentingnya budi pekerti di atas penampilan fisik, kesabaran dalam menghadapi cobaan, serta keberanian dan integritas moral. Nilai-nilai ini diajarkan melalui kisah Pangeran Ande-Ande Lumut yang mencari jodoh dan memilih Kleting Kuning yang berhati mulia meskipun berpenampilan sederhana, serta penolakan Kleting Kuning terhadap godaan Yuyu Kangkang. Cerita ini juga terus diwariskan melalui berbagai media, dari tradisi lisan hingga adaptasi modern, menjadikannya jembatan antar generasi yang melestarikan kearifan lokal.	Cerita Ande-Ande Lumut relevan karena mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas.	chunk-4000, chunk-5000, chunk-5500
Q11	Dokumen tidak menyebutkan siapa pengarang asli cerita Ande-Ande Lumut. Cerita ini merupakan bagian dari warisan budaya takbenda yang diwariskan secara lisan dari generasi ke generasi.		-
Q12	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	null	null
Q13	Cerita Ande-Ande Lumut merupakan sebuah mahakarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa. Kisah ini telah melampaui batasan waktu dan ruang, membuktikan kekuatan narasi lisan dalam membentuk moral, etika, dan identitas sebuah bangsa. Tidak ditemukan informasi spesifik yang mengaitkan asal-usul cerita ini secara langsung dengan Kerajaan Majapahit.	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	chunk-9500
Q14	Dokumen ini tidak menyebutkan jumlah pasti pementasan Ande-Ande Lumut di Jawa. Namun, disebutkan bahwa kisah ini sering diadaptasi dan dipentaskan dalam seni pertunjukan tradisional seperti ketoprak dan wayang orang di Jawa.	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	chunk-9500

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
Q15	Dokumen tidak menyebutkan apakah kisah Ande-Ande Lumut pernah difilmkan. Namun, disebutkan bahwa kisah ini telah diadaptasi ke dalam bentuk animasi, film pendek, atau audiodrama di platform digital seperti YouTube.	Kisah Ande-Ande Lumut telah diadaptasi ke dalam bentuk animasi, film pendek, atau audiodrama di platform digital seperti YouTube.	chunk-7500
Q16	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena mereka memiliki moralitas yang rendah, terbukti dari kesediaan mereka mencium Yuyu Kangkang demi menyeberangi sungai. Selain itu, aroma amis Yuyu Kangkang masih melekat pada diri mereka, yang menunjukkan bahwa mereka tidak memiliki kemurnian hati yang dicari oleh Ande-Ande Lumut.	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena aroma amis Yuyu Kangkang yang melekat pada diri mereka.	chunk-3000
Q17	Sepanjang perjalanan menuju sayembara Ande-Ande Lumut, Kleting Kuning menunjukkan sikap sabar, tabah, berani, dan berintegritas. Meskipun hidup dalam penindasan dan penderitaan akibat perlakuan ibu tiri serta saudara-saudara tirinya, ia tidak pernah membalas dendam atau menyerah pada nasib. Keberaniannya terlihat jelas saat ia menolak mentah-mentah permintaan Yuyu Kangkang untuk dicium demi menyeberang sungai, dan malah menendangnya hingga tak berdaya. Sikap ini mencerminkan keteguhan prinsipnya untuk tidak mengorbankan kehormatan dan harga diri demi kemudahan sesaat.	Kleting Kuning menunjukkan keberanian dan harga diri yang tinggi, menolak permintaan kotor Yuyu Kangkang dan menendangnya.	chunk-2500
Q18	Yuyu Kangkang adalah sosok mitologis berupa kepiting raksasa yang menjaga sungai dalam cerita Ande-Ande Lumut. Peran utamanya adalah sebagai rintangan krusial yang menguji moralitas para Kleting yang ingin menyeberang. Ia menawarkan jasa penyeberangan dengan imbalan ciuman, yang diterima oleh Kleting Merah, Biru, dan Hijau, menunjukkan rendahnya nilai moral mereka. Namun, Kleting Kuning menolak permintaan tersebut dengan tegas, menendang Yuyu Kangkang, dan menyeberang dengan kekuatannya sendiri, yang	Yuyu Kangkang adalah sosok mitologis yang menjaga sungai dan menawarkan jasa penyeberangan dengan imbalan ciuman dari setiap gadis.	chunk-2500

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
	mencerminkan keteguhan prinsip dan integritasnya.		
Q19	Dalam cerita Ande-Ande Lumut, ujian sungai digambarkan sebagai rintangan krusial yang dijaga oleh Yuyu Kangkang, seekor kepiting raksasa. Para Kleting lainnya rela mencium Yuyu Kangkang demi menyeberang, menunjukkan rendahnya moralitas mereka. Namun, Kleting Kuning menolak dan justru menendang Yuyu Kangkang, menunjukkan keteguhan prinsipnya.	Perjalanan para Kleting menuju sayembara Ande-Ande Lumut melibatkan sebuah rintangan krusial: penyeberangan sungai yang dijaga oleh sosok mitologis bernama Yuyu Kangkang. Yuyu Kangkang, seekor kepiting raksasa penjaga sungai, menawarkan jasa penyeberangan dengan imbalan ciuman dari setiap gadis.	chunk-2500
Q20	Kleting Kuning berbeda dari saudara-saudaranya karena ia memiliki hati yang tulus, sabar, dan takwa, meskipun hidup dalam kemiskinan dan penindasan. Berbeda dengan Kleting Merah, Kleting Biru, dan Kleting Hijau yang seringkali digambarkan mementingkan penampilan dan rela mengorbankan moralitas demi kemudahan, Kleting Kuning menunjukkan keberanian dan integritas dengan menolak permintaan Yuyu Kangkang yang tidak pantas, serta menyeberangi sungai dengan kekuatannya sendiri. Akhirnya, kebaikan hati dan kemurnian jiwanya yang diakui oleh Pangeran Ande-Ande Lumut.	Kleting Kuning adalah antitesis dari para gadis pelamar lainnya. Ia digambarkan sebagai gadis yang miskin, lusuh, dan seringkali kotor karena pekerjaan kasar yang dibebankan oleh ibu tiri dan saudara-saudara tirinya—Kleting Merah, Kleting Biru, dan Kleting Hijau—yang kejam.	chunk-2000
Q21	Kisah Ande-Ande Lumut mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas dalam memegang teguh prinsip moral. Cerita ini menekankan bahwa nilai seseorang terletak pada karakter, kejujuran, dan kebaikan hati, bukan pada kemewahan atau kecantikan semata.	Kisah Ande-Ande Lumut mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas dalam memegang prinsip.	chunk-4000, chunk-5000, chunk-5500
Q22	Nilai yang paling menonjol dalam cerita Ande-Ande Lumut adalah penekanan pada budi pekerti dan kebaikan hati di atas penampilan fisik. Cerita ini mengajarkan bahwa ketulusan, kesabaran, dan integritas seseorang jauh lebih berharga daripada kecantikan lahiriah atau status sosial. Hal ini dicontohkan melalui pemilihan Kleting Kuning oleh Pangeran Ande-Ande Lumut,	Cerita ini mengajarkan prioritas budi pekerti di atas penampilan fisik, kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas.	chunk-4000, chunk-5000, chunk-5500

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
	yang meskipun berpenampilan lusuh, memiliki hati yang mulia dan menolak kompromi moral, berbeda dengan saudara-saudara tirinya yang lebih mementingkan penampilan.		
Q23	Kisah Ande-Ande Lumut memiliki makna budaya yang mendalam, mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas dalam memegang prinsip moral. Cerita ini menjadi cerminan nilai-nilai luhur dan instrumen pendidikan karakter dalam budaya Jawa, yang terus diwariskan melalui tradisi lisan dan seni pertunjukan.	Di balik alur ceritanya yang sederhana, Ande-Ande Lumut mengandung lapisan makna dan pesan moral yang sangat dalam, menjadikannya instrumen efektif dalam pendidikan karakter dan pewarisan nilai-nilai luhur.	chunk-4000
Q24	Pesan utama perjalanan Kleting Kuning adalah penekanan pada nilai-nilai luhur seperti budi pekerti, kesabaran, ketabahan, keberanian, dan integritas di atas penampilan fisik atau kemewahan. Cerita ini mengajarkan bahwa kebaikan hati dan kemurnian jiwa akan mendapatkan penghargaan yang setimpal, bahkan dalam menghadapi kesulitan dan penindasan.	Perjalanan Kleting Kuning mengajarkan pentingnya kesabaran, ketabahan dalam menghadapi cobaan, dan keberanian memegang teguh prinsip.	chunk-5000, chunk-5500
Q25	Cerita Ande-Ande Lumut tetap relevan karena mengandung pesan moral yang mendalam dan universal, seperti pentingnya budi pekerti di atas penampilan fisik, kesabaran dalam menghadapi cobaan, serta keberanian dan integritas moral. Nilai-nilai ini diajarkan melalui kisah Pangeran Ande-Ande Lumut yang mencari jodoh dan memilih Kleting Kuning yang berhati mulia meskipun berpenampilan sederhana, serta penolakan Kleting Kuning terhadap godaan Yuyu Kangkang. Cerita ini terus diwariskan melalui berbagai media, dari tradisi lisan hingga adaptasi modern, menjadikannya jembatan antar generasi yang mengajarkan kearifan lokal.	Cerita Ande-Ande Lumut tetap relevan karena mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas.	chunk-4000, chunk-5000, chunk-5500
Q26	Dokumen tersebut tidak menyebutkan siapa pengarang asli dari cerita Ande-Ande Lumut. Cerita ini lebih dijelaskan sebagai sebuah mahakarya takbenda yang berdenyut	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	chunk-9500

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
	dalam nadi kebudayaan Indonesia, khususnya Jawa, dan telah melampaui batasan waktu dan ruang sebagai narasi lisan.		
Q27	Informasi mengenai tahun pertama kali cerita Ande-Ande Lumut ditulis tidak tersedia dalam dokumen yang diberikan. Dokumen tersebut menjelaskan bahwa Ande-Ande Lumut adalah warisan budaya takbenda yang dituturkan secara lisan dan telah melampaui batasan waktu.	Maaf saya belum mendapatkan pelatihan tentang pertanyaan yang anda berikan, silahkan ajukan pertanyaan lainnya.	chunk-9500
Q28	Berdasarkan dokumen yang diberikan, tidak disebutkan bahwa cerita Ande-Ande Lumut berasal dari Kerajaan Majapahit. Cerita ini dijelaskan sebagai mahakarya takbenda yang berdenyut dalam kebudayaan Indonesia, khususnya Jawa.	Maaf, informasi mengenai asal cerita ini dari Kerajaan Majapahit tidak ditemukan dalam dokumen yang diberikan.	chunk-9500
Q29	Jumlah pementasan Ande-Ande Lumut di Jawa tidak disebutkan secara spesifik dalam dokumen. Namun, disebutkan bahwa kisah ini sering diadaptasi dan dipentaskan dalam ketoprak dan wayang orang, serta oleh berbagai sanggar seni dan komunitas budaya.	Maaf, informasi mengenai jumlah pementasan Ande-Ande Lumut di Jawa tidak ditemukan dalam dokumen yang diberikan.	chunk-9500
Q30	Kisah Ande-Ande Lumut berkontribusi dalam pembentukan nilai moral masyarakat Jawa dengan mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas. Cerita ini berfungsi sebagai instrumen pendidikan karakter dan pewarisan nilai-nilai luhur.	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya !	chunk-9500
Q31	Yuyu Kangkang melambangkan ujian moral atau godaan yang menguji integritas dan prinsip para Kleting. Penolakan Kleting Kuning terhadap Yuyu Kangkang mencerminkan keberanian dan keteguhan dalam mempertahankan kehormatan serta nilai moral.	Yuyu Kangkang adalah simbol kuat penolakan Kleting Kuning terhadap kompromi moral, menekankan pentingnya memegang teguh prinsip kebenaran dan kejujuran.	chunk-5500
Q32	Kesabaran dan ketabahan Kleting Kuning dianggap sebagai teladan karena ia menghadapi perlakuan kejam dari ibu tiri dan saudara tirinya tanpa membalas dendam atau	Kleting Kuning adalah teladan kesabaran dan ketabahan karena menghadapi perlakuan kejam tanpa membalas dendam, serta	chunk-4500, chunk-5000

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
	menyerah. Ketabahannya dalam menanggung penderitaan serta menghadapi rintangan seperti Yuyu Kangkang pada akhirnya membuahkan hasil yang adil dan bermakna.	rintangan Yuyu Kangkang, yang membuahkan hasil manis.	
Q33			
Q34	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	null	null
Q35	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	null	null
Q36	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	null	null
Q37	Penolakan Kleting Kuning terhadap Yuyu Kangkang mencerminkan keberanian, integritas, dan keteguhan prinsip. Ia menolak kompromi moral, tidak menukar kehormatan dan harga diri demi kemudahan sesaat, serta memegang teguh kebenaran dan kejujuran.	Penolakan Kleting Kuning terhadap Yuyu Kangkang menunjukkan keberanian, integritas, dan penolakan kompromi moral. Ia tidak menukar kehormatan demi kemudahan sesaat, memegang teguh prinsip kebenaran dan kejujuran.	chunk-2500, chunk-3000, chunk-5000, chunk-5500
Q38	Kleting Kuning digambarkan sebagai sosok yang sangat sabar. Ia menghadapi perlakuan kejam dari ibu dan saudara tirinya tanpa membalas dendam atau menyerah. Kesabaran dan ketabahannya dalam menanggung penderitaan, termasuk rintangan Yuyu Kangkang, pada akhirnya membuahkan hasil manis.	Kleting Kuning digambarkan sebagai figur yang sabar dan tabah dalam menghadapi cobaan, termasuk perlakuan kejam dari ibu tiri dan rintangan Yuyu Kangkang. Kesabarannya membuahkan hasil manis.	chunk-4500, chunk-5000
Q39	Kisah Ande-Ande Lumut diwariskan dalam masyarakat Jawa melalui penuturan lisan dan berbagai bentuk adaptasi, mulai dari tradisi lisan hingga seni pertunjukan dan media modern.	Kisah Ande-Ande Lumut diwariskan secara lisan di balai desa, teras rumah saat senja, atau sebagai pengantar tidur.	chunk-6500
Q40	Ande-Ande Lumut dianggap penting sebagai mahakarya budaya takbenda yang membentuk moral, etika, dan identitas bangsa melalui kekuatan narasi lisan.	Ande-Ande Lumut adalah mahakarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa, membuktikan kekuatan narasi lisan dalam membentuk moral, etika, dan identitas.	chunk-500, chunk-1000
Q41	Kisah Ande-Ande Lumut dapat digunakan untuk pembelajaran karakter karena mengandung nilai budi pekerti, kesabaran, ketabahan, keberanian, dan integritas yang dicontohkan melalui tokoh Kleting Kuning.	Kleting Kuning digambarkan sebagai figur yang sabar dan tabah dalam menghadapi perlakuan kejam tanpa membalas dendam.	chunk-5000

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
Q42	Cerita Ande-Ande Lumut relevan karena mengajarkan prioritas budi pekerti di atas penampilan fisik, kesabaran dan ketabahan, serta keberanian dan integritas dalam menolak kompromi moral.	Cerita Ande-Ande Lumut mengajarkan prioritas budi pekerti di atas penampilan fisik, kesabaran dalam menghadapi cobaan, dan keberanian memegang teguh prinsip moral.	chunk-4000, chunk-4500, chunk-5000, chunk-5500
Q43	Cerita rakyat seperti Ande-Ande Lumut berperan sebagai media pewarisan nilai luhur dan pembentuk identitas budaya melalui kekuatan narasi lisan.	Ande-Ande Lumut adalah mahakarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, membuktikan kekuatan narasi lisan dalam membentuk moral, etika, dan identitas bangsa.	chunk-1000, chunk-4000, chunk-5000, chunk-5500
Q44	Nilai kesabaran Kleting Kuning penting karena menunjukkan ketekunan dan ketabahan dalam menghadapi perlakuan buruk dan rintangan tanpa membalas dendam hingga akhirnya membuahkan hasil positif.	Kleting Kuning digambarkan sebagai figur yang sabar dan tabah dalam menghadapi perlakuan kejam dan rintangan seperti Yuyu Kangkang hingga memperoleh hasil yang baik.	chunk-5000
Q45	Adaptasi digital membantu memperluas jangkauan dan menjaga relevansi cerita di kalangan generasi yang akrab dengan teknologi.	Adaptasi digital cerita Ande-Ande Lumut dalam bentuk animasi, film pendek, atau audiodrama menjangkau audiens yang lebih luas dan akrab dengan teknologi.	chunk-7500, chunk-8000
Q46	Kisah Ande-Ande Lumut memiliki makna penting sebagai cerminan nilai-nilai luhur budaya Jawa, berfungsi sebagai instrumen pendidikan karakter dan pewarisan nilai-nilai luhur antar generasi.	Kisah Ande-Ande Lumut mengandung lapisan makna dan pesan moral yang sangat dalam, menjadikannya instrumen efektif dalam pendidikan karakter dan pewarisan nilai-nilai luhur.	chunk-4000
Q47	Kisah Ande-Ande Lumut dipahami sebagai cerita bermuatan nilai luhur yang menekankan budi pekerti, kesabaran, ketabahan, serta keberanian dan integritas moral.	Di balik alur ceritanya yang sederhana, Ande-Ande Lumut mengandung lapisan makna dan pesan moral yang sangat dalam, menjadikannya instrumen efektif dalam pendidikan karakter dan pewarisan nilai-nilai luhur.	chunk-4000, chunk-5000, chunk-5500
Q48	Tokoh Kleting Kuning mengajarkan prioritas budi pekerti di atas penampilan fisik, kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas dalam memegang prinsip moral.	Kleting Kuning mengajarkan prioritas budi pekerti di atas penampilan fisik, kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas.	chunk-4000, chunk-4500, chunk-5000, chunk-5500

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
Q49	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!	null	null
Q50	Pesan utama cerita Ande-Ande Lumut adalah pentingnya budi pekerti, kesabaran, ketabahan, serta keberanian dan integritas di atas penampilan fisik.	Pesan utama cerita Ande-Ande Lumut adalah prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan, serta keberanian dan integritas.	chunk-4000, chunk-4500, chunk-5000, chunk-5500
Q51	Dokumen tidak menyebutkan secara spesifik masa kerajaan tempat kisah Ande-Ande Lumut pertama kali berkembang di Jawa.	Keterangan bahwa informasi masa kerajaan tidak ditemukan dalam dokumen.	chunk-9500
Q52	Dokumen yang diberikan tidak menyebutkan kerajaan asal Ande-Ande Lumut secara spesifik. Ia digambarkan sebagai seorang putra mahkota.	Maaf saya belum mendapatkan pelatihan tentang pertanyaan yang anda berikan, silahkan ajukan pertanyaan lainnya.	chunk-9500
Q53	Kisah Ande-Ande Lumut tidak memiliki latar waktu sejarah tertentu yang dapat diidentifikasi secara kronologis dan digambarkan sebagai dongeng yang melampaui batasan waktu dan ruang.	Kisah Ande-Ande Lumut telah melampaui batasan waktu dan ruang, tidak memiliki latar waktu sejarah tertentu.	chunk-500
Q54	Yuyu Kangkang digambarkan sebagai sosok mitologis berupa kepiting raksasa penjaga sungai, namun dokumen tidak menjelaskan inspirasi mitologis spesifik dalam kepercayaan Jawa kuno.	Informasi mengenai inspirasi mitologis Yuyu Kangkang tidak ditemukan dalam dokumen.	chunk-9500
Q55	Tidak ada informasi dalam dokumen yang menyatakan penggunaan resmi cerita Ande-Ande Lumut sebagai alat propaganda budaya oleh pemerintah.	Informasi mengenai penggunaan cerita Ande-Ande Lumut sebagai alat propaganda budaya oleh pemerintah tidak ditemukan dalam dokumen.	chunk-9500
Q56	Informasi mengenai tokoh sejarawan Jawa yang pertama kali mendokumentasikan kisah Ande-Ande Lumut dalam manuskrip tertulis tidak ditemukan dalam dokumen yang diberikan.	Maaf saya belum mendapatkan pelatihan tentang pertanyaan yang anda berikan, silahkan ajukan pertanyaan lainnya.	chunk-9500
Q57	Informasi mengenai tahun dan lokasi ditemukannya naskah kuno Ande-Ande Lumut tidak ditemukan dalam dokumen yang diberikan. Cerita ini lebih dikenal sebagai warisan budaya takbenda dan narasi lisan yang diwariskan secara turun-temurun.	Maaf saya belum mendapatkan pelatihan tentang pertanyaan yang anda berikan, silahkan ajukan pertanyaan lainnya.	chunk-9500
Q58	Dokumen yang diberikan tidak menjelaskan hubungan genealogis Ande-Ande Lumut dengan tokoh-tokoh kerajaan Jawa lainnya menurut	Maaf saya belum mendapatkan pelatihan tentang pertanyaan yang anda berikan, silahkan ajukan pertanyaan lainnya.	chunk-9500

Query_ID	Candidate_Text (Indonesian)	Reference_Text (Indonesian)	Chunk_ID
	catatan sejarah. Ande-Ande Lumut dijelaskan sebagai dongeng dan mahakarya takbenda.		
Q59	Berdasarkan dokumen yang diberikan, tidak ada informasi mengenai ritual budaya khusus yang terinspirasi dari kisah Yuyu Kangkang dalam tradisi Jawa kuno. Yuyu Kangkang disebutkan sebagai sosok mitologis dalam cerita Ande-Ande Lumut.	Maaf saya belum mendapatkan pelatihan tentang pertanyaan yang anda berikan, silahkan ajukan pertanyaan lainnya.	chunk-9500
Q60	Dokumen ini tidak menyebutkan versi terpendek atau terpanjang dari cerita Ande-Ande Lumut. Dokumen hanya membahas esensi cerita dan adaptasinya dalam berbagai bentuk.	Maaf, dokumen ini tidak menyebutkan versi terpendek atau terpanjang dari cerita Ande-Ande Lumut.	chunk-9500
Q61	Para Kleting tidak menyeberang sungai menggunakan speedboat karena cerita Ande-Ande Lumut adalah dongeng tradisional yang berlatar waktu sebelum teknologi modern seperti speedboat ada. Penyeberangan sungai dalam cerita tersebut melibatkan sosok mitologis Yuyu Kangkang.	Para Kleting menyeberang sungai yang dijaga Yuyu Kangkang. Kleting Kuning menyeberang dengan kekuatannya sendiri.	chunk-2500, chunk-3000

Appendix E: The Completed Automatic Metric Schema

Table 6 – The Completed Automatic Metric Schema

Query_ID	Automatic Metric Data (BERTScore (F1), ROUGE-L (F1), Cosine Similarity, Similarity_Label, Hallucination)
Q1	{'BERTScore (F1)': 0.8878, 'ROUGE-L (F1)': 0.6667, 'Cosine_Similarity': np.float32(0.9415), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q2	{'BERTScore (F1)': 0.8335, 'ROUGE-L (F1)': 0.4848, 'Cosine_Similarity': np.float32(0.8809), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q3	{'BERTScore (F1)': 0.7661, 'ROUGE-L (F1)': 0.254, 'Cosine_Similarity': np.float32(0.7024), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q4	{'BERTScore (F1)': 0.805, 'ROUGE-L (F1)': 0.4375, 'Cosine_Similarity': np.float32(0.9036), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Entity hallucination'}
Q5	{'BERTScore (F1)': 0.8247, 'ROUGE-L (F1)': 0.4615, 'Cosine_Similarity': np.float32(0.8733), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Entity hallucination'}
Q6	{'BERTScore (F1)': 0.8637, 'ROUGE-L (F1)': 0.5373, 'Cosine_Similarity': np.float32(0.7845), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q7	{'BERTScore (F1)': 0.811, 'ROUGE-L (F1)': 0.3902, 'Cosine_Similarity': np.float32(0.8572), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Entity hallucination'}
Q8	{'BERTScore (F1)': 0.7359, 'ROUGE-L (F1)': 0.1481, 'Cosine_Similarity': np.float32(0.5736), 'Similarity_Label': 'Rendah', 'Hallucination': 'Potensi halusinasi'}
Q9	{'BERTScore (F1)': 0.7806, 'ROUGE-L (F1)': 0.2192, 'Cosine_Similarity': np.float32(0.7164), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q10	{'BERTScore (F1)': 0.7909, 'ROUGE-L (F1)': 0.367, 'Cosine_Similarity': np.float32(0.7094), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q11	n/a
Q12	n/a
Q13	n/a
Q14	n/a
Q15	{'BERTScore (F1)': 0.8891, 'ROUGE-L (F1)': 0.7755, 'Cosine_Similarity': np.float32(0.861), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q16	{'BERTScore (F1)': 0.8509, 'ROUGE-L (F1)': 0.5217, 'Cosine_Similarity': np.float32(0.8163), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q17	{'BERTScore (F1)': 0.7462, 'ROUGE-L (F1)': 0.2174, 'Cosine_Similarity': np.float32(0.6979), 'Similarity_Label': 'Rendah', 'Hallucination': 'Entity hallucination'}
Q18	{'BERTScore (F1)': 0.7821, 'ROUGE-L (F1)': 0.3111, 'Cosine_Similarity': np.float32(0.7269), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q19	{'BERTScore (F1)': 0.7662, 'ROUGE-L (F1)': 0.3171, 'Cosine_Similarity': np.float32(0.8286), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q20	{'BERTScore (F1)': 0.7484, 'ROUGE-L (F1)': 0.2632, 'Cosine_Similarity': np.float32(0.7298), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q21	{'BERTScore (F1)': 0.9013, 'ROUGE-L (F1)': 0.7123, 'Cosine_Similarity': np.float32(0.8199), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}

Query_ID	Automatic Metric Data (BERTScore (F1), ROUGE-L (F1), Cosine Similarity, Similarity_Label, Hallucination)
Q22	{'BERTScore (F1)': 0.7467, 'ROUGE-L (F1)': 0.2174, 'Cosine_Similarity': np.float32(0.6557), 'Similarity_Label': 'Rendah', 'Hallucination': 'Entity hallucination'}
Q23	{'BERTScore (F1)': 0.7672, 'ROUGE-L (F1)': 0.2381, 'Cosine_Similarity': np.float32(0.8319), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q24	{'BERTScore (F1)': 0.7888, 'ROUGE-L (F1)': 0.2667, 'Cosine_Similarity': np.float32(0.7919), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q25	{'BERTScore (F1)': 0.7978, 'ROUGE-L (F1)': 0.3853, 'Cosine_Similarity': np.float32(0.7014), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q26	n/a
Q27	n/a
Q28	n/a
Q29	n/a
Q30	n/a
Q31	{'BERTScore (F1)': 0.797, 'ROUGE-L (F1)': 0.2857, 'Cosine_Similarity': np.float32(0.8904), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q32	{'BERTScore (F1)': 0.8297, 'ROUGE-L (F1)': 0.5, 'Cosine_Similarity': np.float32(0.8821), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q33	n/a
Q34	n/a
Q35	n/a
Q36	{'BERTScore (F1)': 0.9372, 'ROUGE-L (F1)': 0.7719, 'Cosine_Similarity': np.float32(0.9571), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q37	{'BERTScore (F1)': 0.8623, 'ROUGE-L (F1)': 0.5714, 'Cosine_Similarity': np.float32(0.9198), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q38	{'BERTScore (F1)': 0.7904, 'ROUGE-L (F1)': 0.2791, 'Cosine_Similarity': np.float32(0.5982), 'Similarity_Label': 'Rendah', 'Hallucination': 'Potensi halusinasi'}
Q39	{'BERTScore (F1)': 0.8712, 'ROUGE-L (F1)': 0.5, 'Cosine_Similarity': np.float32(0.8885), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q40	{'BERTScore (F1)': 0.7352, 'ROUGE-L (F1)': 0.0976, 'Cosine_Similarity': np.float32(0.751), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q41	{'BERTScore (F1)': 0.9023, 'ROUGE-L (F1)': 0.6809, 'Cosine_Similarity': np.float32(0.9623), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q42	{'BERTScore (F1)': 0.7942, 'ROUGE-L (F1)': 0.2791, 'Cosine_Similarity': np.float32(0.6127), 'Similarity_Label': 'Rendah', 'Hallucination': 'Potensi halusinasi'}
Q43	{'BERTScore (F1)': 0.7967, 'ROUGE-L (F1)': 0.4255, 'Cosine_Similarity': np.float32(0.786), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q44	{'BERTScore (F1)': 0.7442, 'ROUGE-L (F1)': 0.3684, 'Cosine_Similarity': np.float32(0.7131), 'Similarity_Label': 'Sedang', 'Hallucination': 'Borderline'}
Q45	{'BERTScore (F1)': 0.87, 'ROUGE-L (F1)': 0.52, 'Cosine_Similarity': np.float32(0.6367), 'Similarity_Label': 'Rendah', 'Hallucination': 'Borderline'}
Q46	{'BERTScore (F1)': 0.7693, 'ROUGE-L (F1)': 0.2, 'Cosine_Similarity': np.float32(0.8048), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q47	{'BERTScore (F1)': 0.9609, 'ROUGE-L (F1)': 0.8889, 'Cosine_Similarity': np.float32(0.8034), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q48	n/a
Q49	{'BERTScore (F1)': 0.9545, 'ROUGE-L (F1)': 0.7143, 'Cosine_Similarity': np.float32(0.973), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}

Query_ID	Automatic Metric Data (BERTScore (F1), ROUGE-L (F1), Cosine Similarity, Similarity_Label, Hallucination)
Q50	{'BERTScore (F1)': 0.755, 'ROUGE-L (F1)': 0.1538, 'Cosine_Similarity': np.float32(0.7078), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}
Q51	n/a
Q52	{'BERTScore (F1)': 0.87, 'ROUGE-L (F1)': 0.4878, 'Cosine_Similarity': np.float32(0.8037), 'Similarity_Label': 'Sedang', 'Hallucination': 'Tidak terindikasi'}
Q53	n/a
Q54	{'BERTScore (F1)': 0.9253, 'ROUGE-L (F1)': 0.6667, 'Cosine_Similarity': np.float32(0.9648), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q55	n/a
Q56	n/a
Q57	n/a
Q58	n/a
Q59	{'BERTScore (F1)': 0.8973, 'ROUGE-L (F1)': 0.7027, 'Cosine_Similarity': np.float32(0.8982), 'Similarity_Label': 'Tinggi', 'Hallucination': 'Tidak terindikasi'}
Q60	{'BERTScore (F1)': 0.7551, 'ROUGE-L (F1)': 0.2917, 'Cosine_Similarity': np.float32(0.6295), 'Similarity_Label': 'Rendah', 'Hallucination': 'Entity hallucination'}
Q61	{'BERTScore (F1)': 0.762, 'ROUGE-L (F1)': 0.2553, 'Cosine_Similarity': np.float32(0.821), 'Similarity_Label': 'Sedang', 'Hallucination': 'Entity hallucination'}

Appendix F: Human-in-the-Loop (HITL) Annotation Schema

Table 7 - Human-in-the-Loop (HITL) Annotation Schema

Field Name	Data Type	Description	Source / Stage
query_id	String	Unique identifier of the evaluated query. Serves as the primary alignment anchor for all evaluation artifacts.	Query Design
query_text	Text	The question submitted to both the RAG system and the LLM Judge.	Query Input
query_category	Categorical	Context category of the query (in-context, ambiguous, out-of-context).	Query Design
candidate_response	Text	Response generated by the LLM RAG, representing the system output under evaluation.	LLM RAG
reference_response	Text	Source-based reference response generated by the LLM Judge using retrieved chunks.	LLM Judge
chunk_id	String	Identifier of the chunk referenced as the grounding basis for the response.	LLM Judge
llm_judge_status	Boolean	Logical status of the LLM Judge output (true = valid output, false = system-level failure).	LLM Judge
bert_score_f1	Float	Semantic similarity score computed using BERTScore (F1).	Automatic Screening
rouge_l	Float	Text overlap-based similarity score computed using BERTSCORE (F1).	Automatic Screening
cosine_similarity	Float	Vector similarity score between candidate and reference responses.	Automatic Screening
similarity_label	Categorical	Coarse-grained similarity label (e.g., high, medium, low).	Automatic Screening
hallucination_flag	Boolean	Initial indication of potential hallucination based on automatic screening signals.	Automatic Screening
hitl_decision	Categorical	Final decision made by human evaluators (Accept, Abstain, Failure).	HITL
error_code	Categorical	HITL error code (E0-E10) as defined in Table 8.	HITL
failure_type	Categorical	Type of failure identified (Content or System).	HITL
evaluator_id	String	Identifier of the HITL evaluator who provided the annotation.	HITL
annotation_note	Text	Optional qualitative note provided by the evaluator to justify or clarify the decision.	HITL

Appendix G: Distribution of Evaluation Scores per Evaluator

Distribution of Evaluation Scores per Evaluator

Score Category	Evaluator 1	Evaluator 2	Evaluator 3
E0 (Fully Grounded)	18 (29.5%)	15 (24.6%)	20 (32.8%)
E1-E4 (Minor to Moderate Content Issues)	32 (52.5%)	28 (45.9%)	30 (49.2%)
E5-E6 (Ambiguous or Unsupported Content)	2 (3.3%)	3 (4.9%)	1 (1.6%)
E7-E8 (Hallucination or Contradiction)	1 (1.6%)	2 (3.3%)	1 (1.6%)
E9 (Abstain / No Answer)	5 (8.2%)	8 (13.1%)	6 (9.8%)
E10 (System Failure)	3 (4.9%)	5 (8.2%)	3 (4.9%)

Appendix H : Pairwise Inter-Evaluator Agreement Based on Cohen's Kappa

Pairwise Inter-Evaluator Agreement Based on Cohen's Kappa

Evaluator Pair	Cohen's Kappa	Agreement Level
Evaluator 1 - Evaluator 2	0.72	Substantial
Evaluator 1 - Evaluator 3	0.68	Substantial
Evaluator 2 - Evaluator 3	0.65	Substantial

Cohen's Kappa values were computed using the initial independent judgments of each evaluator prior to consensus, based on binary outcome categories (Accepted / Rejected).

Pairwise inter-evaluator agreement was assessed using Cohen's Kappa to measure consistency between individual evaluators prior to consensus. All evaluator pairs exhibited substantial agreement, indicating a stable and reliable multi-evaluator annotation process.

Appendix I: Multi-Evaluator Agreement Using Fleiss' Kappa

Multi-Evaluator Agreement Using Fleiss' Kappa

Number of Evaluators	Fleiss' Kappa	Agreement Level
3	0.69	Substantial

Fleiss' Kappa was calculated based on the independent initial evaluations of three human evaluators before consensus reconciliation, using binary classification outcomes.

Fleiss' Kappa was employed to assess overall agreement among three evaluators simultaneously. The resulting substantial agreement score further supports the internal consistency of the HITL evaluation framework

Appendix J: The Full Set HITL Evaluation Results

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
Q1	E0	E4	E2	Accepted	The core answer is supported by the chunk (the foul smell of Yuyu Kangkang), but there is an addition of a normative interpretation about sacrificing dignity and integrity that is not explicitly stated in chunk-3000.	No Discussion Needed	Mengapa Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau?	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena mereka memiliki aroma amis Yuyu Kangkang yang melekat pada diri mereka. Ini menunjukkan bahwa mereka telah mengorbankan martabat demi kemudahan menyeberangi sungai, berbeda dengan Kleting Kuning yang berintegritas.
Q2	E2	E0	E1	Accepted	The answer is strongly supported by multiple chunks (chunk-2000, 2500, 5000) and semantically consistent, although the phrase "throughout the journey" is not elaborated chronologically.	No Discussion Needed	Bagaimana sikap Kleting Kuning sepanjang perjalanan?	Sepanjang perjalanan, Kleting Kuning menunjukkan sikap tulus, sabar, dan takwa. Ia memiliki keberanian dan integritas tinggi, menolak mentah-mentah permintaan kotor Yuyu Kangkang dengan menendangnya, serta menyeberangi sungai dengan kekuatannya sendiri. Ia juga tabah menghadapi penderitaan.
Q3	E4	E0	E2	Accepted	Yuyu Kangkang's role as an obstacle and river guardian is fully supported by the chunk. However, the claim that Kleting Kuning's rejection is the "main distinguishing factor in the Prince's eyes" is	No Discussion Needed	Apa peran Yuyu Kangkang dalam alur cerita?	Yuyu Kangkang berperan sebagai rintangan krusial dalam perjalanan para Kleting menuju sayembara Ande-Ande Lumut. Ia menguji moralitas para gadis dengan menawarkan penyeberangan sungai dengan imbalan

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
					interpretative and not explicitly stated in the chunk.			ciuman. Penolakan Kleting Kuning terhadap Yuyu Kangkang menunjukkan keberanian dan integritasnya, yang menjadi pembeda utama di mata Pangeran Ande-Ande Lumut.
Q4	E4	E4	E1	Rejected	The description of the river test is consistent with the chunk, but the action detail of "kicking Yuyu Kangkang" and the evaluative meaning of "showing courage and integrity" are not fully explicated in chunk-2500.	Discussion Needed	Bagaimana ujian sungai digambarkan dalam cerita?	Ujian sungai dalam cerita Ande-Ande Lumut digambarkan sebagai rintangan krusial yang dijaga oleh Yuyu Kangkang, seekor kepiting raksasa. Ia menawarkan penyeberangan dengan imbalan ciuman. Gadis-gadis lain rela menciumnya, namun Kleting Kuning menolak tegas, menendang Yuyu Kangkang, dan menyeberang sendiri, menunjukkan keberanian serta integritasnya.
Q5	E2	E0	E2	Accepted	The differences in attitude and morality of Kleting Kuning are supported by several chunks. However, the explicit statement that Ande-Ande Lumut chose her "based on purity of soul, not physical appearance" is an interpretative synthesis across chunks.	No Discussion Needed	Apa yang membedakan Kleting Kuning dari saudara-saudaranya?	Kleting Kuning berbeda dari saudara-saudaranya karena hati yang tulus, sabar, takwa, dan integritasnya. Ia menolak mengorbankan harga dirinya dengan mencium Yuyu Kangkang, tidak seperti saudara-saudaranya yang rela melakukannya. Ande-Ande Lumut memilihnya berdasarkan kemurnian jiwa, bukan penampilan fisik.

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
Q6	E0	E0	E0	Accepted	The answer is normative but aligns with the ambiguous Query_Type and is fully supported by the elaboration of meaning and moral message explicitly stated in the reference chunks.	No Discussion Needed	Apa makna kisah Ande-Ande Lumut bagi kehidupan manusia?	Kisah Ande-Ande Lumut mengajarkan bahwa budi pekerti, kesabaran, dan integritas jauh lebih berharga daripada penampilan fisik atau kekayaan. Cerita ini menekankan pentingnya karakter, kejujuran, dan kebaikan hati dalam menghadapi cobaan hidup, serta keberanian untuk memegang teguh prinsip moral.
Q7	E2	E4	E1	Accepted	The answer asserts the core values that are explicit in the chunk. However, the addition of other values (patience, fortitude, courage, integrity), although thematically relevant, are not all stated as "most prominent" in the reference chunks.	No Discussion Needed	Nilai apa yang paling menonjol dalam cerita ini?	Nilai paling menonjol dalam cerita Ande-Ande Lumut adalah prioritas budi pekerti di atas penampilan fisik, serta pentingnya kesabaran, ketabahan, keberanian, dan integritas dalam menghadapi cobaan hidup.
Q8	E4	E4	E2	Rejected	The dimensions of moral meaning and noble values are supported by the chunk. However, explicit claims about "intangible cultural heritage", "bridge between generations", and "national identity" are not stated in the reference chunks, thus constituting interpretative additions beyond the source.	Discussion Needed	Bagaimana kisah ini dapat dimaknai secara budaya?	Kisah Ande-Ande Lumut kaya akan makna budaya, terutama dalam menanamkan nilai-nilai luhur. Cerita ini menekankan pentingnya budi pekerti di atas penampilan fisik, mengajarkan bahwa kebaikan hati dan kejujuran lebih berharga daripada kecantikan lahiriah semata. Selain itu, kisah ini mengilustrasikan

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
								nilai kesabaran dan ketabahan dalam menghadapi kesulitan, serta keberanian untuk mempertahankan integritas dan harga diri. Makna-nilai ini menjadikan Ande-Ande Lumut sebagai warisan budaya takbenda yang berfungsi sebagai jembatan antar generasi, menjaga kearifan lokal dan identitas bangsa.
Q9	E2	E4	E2	Accepted	The main moral message regarding patience, fortitude, courage, and integrity is supported by the chunk. However, the claim about "true happiness" and the explicit explanation of Ande-Ande Lumut choosing Kleting Kuning as a direct result of the journey is an interpretative synthesis.	No Discussion Needed	Apa pesan utama perjalanan Kleting Kuning?	Pesan utama dari perjalanan Kleting Kuning adalah penekanan pada nilai-nilai luhur seperti budi pekerti, kesabaran, ketabahan, keberanian, dan integritas di atas penampilan fisik atau kekayaan. Cerita ini mengajarkan bahwa kebaikan hati dan kemurnian jiwa akan selalu dihargai dan pada akhirnya akan membawa kebahagiaan sejati, seperti yang ditunjukkan oleh Pangeran Ande-Ande Lumut yang memilih Kleting Kuning meskipun penampilannya lusuh.
Q10	E2	E4	E2	Accepted	Universal moral values and character messages are supported by the chunk.	No Discussion Needed	Mengapa cerita Ande-Ande Lumut	Cerita Ande-Ande Lumut tetap relevan karena mengandung pesan moral

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
					However, claims about cross-media sustainability, bridge between generations, and preservation of local wisdom are not explicitly stated in the reference chunks, thus constituting interpretative expansion.		tetap relevan hingga kini?	yang mendalam dan universal, seperti pentingnya budi pekerti di atas penampilan fisik, kesabaran dalam menghadapi cobaan, serta keberanian dan integritas moral. Nilai-nilai ini diajarkan melalui kisah Pangeran Ande-Ande Lumut yang mencari jodoh dan memilih Kleting Kuning yang berhati mulia meskipun berpenampilan sederhana, serta penolakan Kleting Kuning terhadap godaan Yuyu Kangkang. Cerita ini juga terus diwariskan melalui berbagai media, dari tradisi lisan hingga adaptasi modern, menjadikannya jembatan antar generasi yang melestarikan kearifan lokal.
Q11	E9	E0	E9	Accepted	The answer explicitly states the inability to answer based on available information, acknowledging that the document does not mention the original author, and provides alternative context appropriate to the document's content (intangible cultural heritage).	No Discussion Needed	Siapa pengarang asli cerita Ande-Ande Lumut menurut dokumen?	Dokumen tidak menyebutkan siapa pengarang asli cerita Ande-Ande Lumut. Cerita ini merupakan bagian dari warisan budaya takbenda yang diwariskan secara lisan dari generasi ke generasi.

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
Q12	E10	E9	E10	Rejected	The system failed to generate an evaluable answer (empty/non-substantive output), thus categorized as a system failure.	Discussion Needed	Pada tahun berapa cerita ini pertama kali ditulis?	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!
Q13	E6	E10	E9	Rejected	The answer essentially states the absence of specific information in the document regarding the origin of Majapahit. Although there is a descriptive introduction that was not requested, the response still functions as an appropriate abstain for an out-of-context query.	Discussion Needed	Apakah cerita ini berasal dari Kerajaan Majapahit?	Cerita Ande-Ande Lumut merupakan sebuah mahakarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa. Kisah ini telah melampaui batasan waktu dan ruang, membuktikan kekuatan narasi lisan dalam membentuk moral, etika, dan identitas sebuah bangsa. Tidak ditemukan informasi spesifik yang mengaitkan asal-usul cerita ini secara langsung dengan Kerajaan Majapahit.
Q14	E9	E10	E9	Accepted	The answer explicitly states the absence of quantitative data in the document. Additional descriptive information is still contextually relevant, and the use of chunk-9500 is appropriate as an indicator that specific information was not found.	No Discussion Needed	Berapa jumlah pementasan Ande-Ande Lumut di Jawa?	Dokumen ini tidak menyebutkan jumlah pasti pementasan Ande-Ande Lumut di Jawa. Namun, disebutkan bahwa kisah ini sering diadaptasi dan dipentaskan dalam seni pertunjukan tradisional seperti ketoprak dan wayang orang di Jawa.
Q15	E8	E7	E8	Rejected	There is an internal contradiction: the answer	Discussion Needed	Apakah kisah ini pernah	Dokumen tidak menyebutkan apakah kisah Ande-Ande

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
					states that the document does not mention filming, yet simultaneously mentions the existence of adaptations in the form of short films. The second information is supported by a chunk, making the first statement contradictory.		difilmkan menurut dokumen?	Lumut pernah difilmkan. Namun, disebutkan bahwa kisah ini telah diadaptasi ke dalam bentuk animasi, film pendek, atau audiodrama di platform digital seperti YouTube.
Q16	E2	E4	E2	Accepted	The reason for rejection based on the foul smell of Yuyu Kangkang is explicitly supported by the chunk. However, the assessment of "low morality" and "purity of heart being sought" are evaluative interpretations not directly stated in the chunk.	No Discussion Needed	Mengapa Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau?	Ande-Ande Lumut menolak Kleting Merah, Biru, dan Hijau karena mereka memiliki moralitas yang rendah, terbukti dari kesediaan mereka mencium Yuyu Kangkang demi menyeberangi sungai. Selain itu, aroma amis Yuyu Kangkang masih melekat pada diri mereka, yang menunjukkan bahwa mereka tidak memiliki kemurnian hati yang dicari oleh Ande-Ande Lumut.
Q17	E2	E4	E3	Rejected	Chunk-2500 only supports the aspect of courage and rejection of Yuyu Kangkang. Descriptions about patience, fortitude in facing oppression, and not seeking revenge are not fully supported by the referenced chunk, resulting in weak source attribution.	Discussion Needed	Bagaimana sikap Kleting Kuning sepanjang perjalanan?	Sepanjang perjalanan menuju sayembara Ande-Ande Lumut, Kleting Kuning menunjukkan sikap sabar, tabah, berani, dan berintegritas. Meskipun hidup dalam penindasan dan penderitaan akibat perlakuan ibu tiri serta saudara-saudara tirinya, ia tidak pernah

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
								membalas dendam atau menyerah pada nasib. Keberaniannya terlihat jelas saat ia menolak mentah-mentah permintaan Yuyu Kangkang untuk dicium demi menyeberang sungai, dan malah menendangnya hingga tak berdaya. Sikap ini mencerminkan keteguhan prinsipnya untuk tidak mengorbankan kehormatan dan harga diri demi kemudahan sesaat.
Q18	E2	E4	E2	Accepted	Yuyu Kangkang's role as a river guardian and crossing obstacle is explicitly supported by the chunk. However, the assessment of the "low moral value" of the other Kletings and the interpretation of integrity as the main function of the test are evaluative elaborations not directly stated in the chunk.	No Discussion Needed	Apa peran Yuyu Kangkang dalam alur cerita?	Yuyu Kangkang adalah sosok mitologis berupa kepiting raksasa yang menjaga sungai dalam cerita Ande-Ande Lumut. Peran utamanya adalah sebagai rintangan krusial yang menguji moralitas para Kleting yang ingin menyeberang. Ia menawarkan jasa penyeberangan dengan imbalan ciuman, yang diterima oleh Kleting Merah, Biru, dan Hijau, menunjukkan rendahnya nilai moral mereka. Namun, Kleting Kuning menolak permintaan tersebut dengan tegas, menendang Yuyu Kangkang, dan menyeberang dengan

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
								kekuatannya sendiri, yang mencerminkan keteguhan prinsip dan integritasnya.
Q19	E2	E4	E3	Rejected	Chunk-2500 supports the description of the river test as an obstacle and crossing mechanism. However, the action detail of "kicking Yuyu Kangkang" and the explicit assessment of "low morality" are not fully supported by the referenced chunk, so source attribution is considered weak.	Discussion Needed	Bagaimana ujian sungai digambarkan dalam cerita?	Dalam cerita Ande-Ande Lumut, ujian sungai digambarkan sebagai rintangan krusial yang dijaga oleh Yuyu Kangkang, seekor kepiting raksasa. Para Kleting lainnya rela mencium Yuyu Kangkang demi menyeberang, menunjukkan rendahnya moralitas mereka. Namun, Kleting Kuning menolak dan justru menendang Yuyu Kangkang, menunjukkan keteguhan prinsipnya.
Q20	E6	E4	E3	Rejected	Chunk-2000 supports the difference in social status and Kleting Kuning's position as the antithesis of her stepsisters. However, descriptions of the rejection of Yuyu Kangkang, moral integrity, and recognition by Ande-Ande Lumut are not supported by the referenced chunk, so source attribution is considered weak.	Discussion Needed	Apa yang membedakan Kleting Kuning dari saudara-saudaranya?	Kleting Kuning berbeda dari saudara-saudaranya karena ia memiliki hati yang tulus, sabar, dan takwa, meskipun hidup dalam kemiskinan dan penindasan. Berbeda dengan Kleting Merah, Kleting Biru, dan Kleting Hijau yang seringkali digambarkan mementingkan penampilan dan rela mengorbankan moralitas demi kemudahan, Kleting Kuning menunjukkan keberanian dan integritas dengan menolak permintaan Yuyu Kangkang yang tidak

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
								pantas, serta menyeberangi sungai dengan kekuatannya sendiri. Akhirnya, kebaikan hati dan kemurnian jiwanya yang diakui oleh Pangeran Ande-Ande Lumut.
Q21	E2	E0	E0	Accepted	The entire content of the answer aligns with the meaning and moral message explicitly stated in the reference chunks. There are no interpretative additions beyond the normative scope of the ambiguous query.	No Discussion Needed	Apa makna kisah Ande-Ande Lumut bagi kehidupan manusia?	Kisah Ande-Ande Lumut mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas dalam memegang teguh prinsip moral. Cerita ini menekankan bahwa nilai seseorang terletak pada karakter, kejujuran, dan kebaikan hati, bukan pada kemewahan atau kecantikan semata.
Q22	E4	E0	E2	Accepted	The core value of prioritizing character over physical appearance is explicitly supported by the chunk. However, the explanatory concrete example of Ande-Ande Lumut choosing Kleting Kuning is not directly stated in the reference chunk, thus constituting an additional elaboration.	No Discussion Needed	Nilai apa yang paling menonjol dalam cerita ini?	Nilai yang paling menonjol dalam cerita Ande-Ande Lumut adalah penekanan pada budi pekerti dan kebaikan hati di atas penampilan fisik. Cerita ini mengajarkan bahwa ketulusan, kesabaran, dan integritas seseorang jauh lebih berharga daripada kecantikan lahiriah atau status sosial. Hal ini dicontohkan melalui

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
								pemilihan Kleting Kuning oleh Pangeran Ande-Ande Lumut, yang meskipun berpenampilan lusuh, memiliki hati yang mulia dan menolak kompromi moral, berbeda dengan saudara-saudara tirinya yang lebih mementingkan penampilan.
Q23	E2	E4	E3	Rejected	Chunk-4000 supports moral meaning and character education function. However, references to "Javanese culture", "oral tradition", and "performing arts" are not covered in the referenced chunk, so source attribution is considered weak.	Discussion Needed	Bagaimana kisah ini dapat dimaknai secara budaya?	Kisah Ande-Ande Lumut memiliki makna budaya yang mendalam, mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas dalam memegang prinsip moral. Cerita ini menjadi cerminan nilai-nilai luhur dan instrumen pendidikan karakter dalam budaya Jawa, yang terus diwariskan melalui tradisi lisan dan seni pertunjukan.
Q24	E2	E4	E2	Accepted	The core moral message about patience, fortitude, courage, and integrity is supported by the chunk. However, the claim that kindness will "receive commensurate appreciation" is a normative generalization	No Discussion Needed	Apa pesan utama perjalanan Kleting Kuning?	Pesan utama perjalanan Kleting Kuning adalah penekanan pada nilai-nilai luhur seperti budi pekerti, kesabaran, ketabahan, keberanian, dan integritas di atas penampilan fisik atau kemewahan. Cerita ini mengajarkan bahwa kebaikan

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
					and not explicitly stated in the chunk.			hati dan kemurnian jiwa akan mendapatkan penghargaan yang setimpal, bahkan dalam menghadapi kesulitan dan penindasan.
Q25	E2	E4	E2	Accepted	The universal moral values that form the basis of the story's relevance are supported by the chunk. However, claims about cross-media sustainability, modern adaptation, and function as a bridge between generations are not explicitly stated in the reference chunks, thus constituting interpretative expansion.	No Discussion Needed	Mengapa cerita Ande-Ande Lumut tetap relevan hingga kini?	Cerita Ande-Ande Lumut tetap relevan karena mengandung pesan moral yang mendalam dan universal, seperti pentingnya budi pekerti di atas penampilan fisik, kesabaran dalam menghadapi cobaan, serta keberanian dan integritas moral. Nilai-nilai ini diajarkan melalui kisah Pangeran Ande-Ande Lumut yang mencari jodoh dan memilih Kleting Kuning yang berhati mulia meskipun berpenampilan sederhana, serta penolakan Kleting Kuning terhadap godaan Yuyu Kangkang. Cerita ini terus diwariskan melalui berbagai media, dari tradisi lisan hingga adaptasi modern, menjadikannya jembatan antar generasi yang mengajarkan kearifan lokal.
Q26	E9	E10	E9	Accepted	The answer explicitly states the absence of information regarding the author in the document. Additional	No Discussion Needed	Siapa pengarang asli cerita Ande-Ande Lumut	Dokumen tersebut tidak menyebutkan siapa pengarang asli dari cerita Ande-Ande Lumut. Cerita ini

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
					descriptive elaboration does not alter the primary function of the answer as an abstain.		menurut dokumen?	lebih dijelaskan sebagai sebuah mahakarya takbenda yang berdenyut dalam nadi kebudayaan Indonesia, khususnya Jawa, dan telah melampaui batasan waktu dan ruang sebagai narasi lisan.
Q27	E0	E10	E9	Accepted	The answer substantively states the absence of information on the year of writing in the document. Although there is a system fallback message, the main response functions as an appropriate abstain for an out-of-context query.	No Discussion Needed	Pada tahun berapa cerita ini pertama kali ditulis?	Informasi mengenai tahun pertama kali cerita Ande-Ande Lumut ditulis tidak tersedia dalam dokumen yang diberikan. Dokumen tersebut menjelaskan bahwa Ande-Ande Lumut adalah warisan budaya takbenda yang dituturkan secara lisan dan telah melampaui batasan waktu.
Q28	E0	E9	E9	Accepted	The answer consistently states the absence of explicit information regarding the origin from Majapahit in the document. There are no speculative claims or additions of external knowledge.	No Discussion Needed	Apakah cerita ini berasal dari Kerajaan Majapahit?	Berdasarkan dokumen yang diberikan, tidak disebutkan bahwa cerita Ande-Ande Lumut berasal dari Kerajaan Majapahit. Cerita ini dijelaskan sebagai mahakarya takbenda yang berdenyut dalam kebudayaan Indonesia, khususnya Jawa.
Q29	E0	E9	E9	Accepted	The answer explicitly states the absence of quantitative data and only refers to qualitative information that indeed exists in the document.	No Discussion Needed	Berapa jumlah pementasan Ande-Ande Lumut di Jawa?	Jumlah pementasan Ande-Ande Lumut di Jawa tidak disebutkan secara spesifik dalam dokumen. Namun, disebutkan bahwa kisah ini sering diadaptasi dan

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
								dipentaskan dalam ketoprak dan wayang orang, serta oleh berbagai sanggar seni dan komunitas budaya.
Q30	E0	E10	E3	Rejected	Substantive answer given, but not supported by relevant system references or chunks in this context.	Discussion Needed	Bagaimana kisah Ande-Ande Lumut berkontribusi dalam pembentukan nilai moral masyarakat Jawa?	Kisah Ande-Ande Lumut berkontribusi dalam pembentukan nilai moral masyarakat Jawa dengan mengajarkan prioritas budi pekerti di atas penampilan fisik, pentingnya kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas. Cerita ini berfungsi sebagai instrumen pendidikan karakter dan pewarisan nilai-nilai luhur.
Q31	E2	E0	E0	Accepted	The answer is consistent with the story context and supported by relevant chunks.	No Discussion Needed	Apa makna simbolik sosok Yuyu Kangkang dalam perjalanan para Kleting menuju sayembara Ande-Ande Lumut?	Yuyu Kangkang melambangkan ujian moral atau godaan yang menguji integritas dan prinsip para Kleting. Penolakan Kleting Kuning terhadap Yuyu Kangkang mencerminkan keberanian dan keteguhan dalam mempertahankan kehormatan serta nilai moral.
Q32	E0	E0	E0	Accepted	The answer is relevant, consistent, and explicitly supported by reference chunks.	No Discussion Needed	Mengapa kesabaran dan ketabahan Kleting Kuning dianggap	Kesabaran dan ketabahan Kleting Kuning dianggap sebagai teladan karena ia menghadapi perlakuan kejam dari ibu tiri dan saudara

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
							sebagai teladan dalam cerita ini?	tirinya tanpa membalas dendam atau menyerah. Ketabahannya dalam menanggung penderitaan serta menghadapi rintangan seperti Yuyu Kangkang pada akhirnya membuahkan hasil yang adil dan bermakna.
Q33	E10	E10	E9	Rejected	There is no substantive answer nor chunk references in the document. The system explicitly does not provide a response, so evaluators set an abstain even though the question is labeled in-context.	Discussion Needed	Bagaimana adaptasi modern cerita Ande-Ande Lumut membantu menjaga relevansinya di era digital?	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!
Q34	E10	E10	E9	Rejected	There is no answer, references, or supporting chunks in the document. The system explicitly states it does not provide a response, so evaluation sets an abstain.	Discussion Needed	Mengapa Ande-Ande Lumut disebut sebagai jembatan antar generasi dalam konteks warisan budaya takbenda?	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!
Q35	E10	E10	E9	Rejected	No document-based answer, references, or supporting chunks are available. The system provides a fallback without substantive content, so evaluation sets an abstain.	Discussion Needed	Mengapa Ande-Ande Lumut memilih Kleting Kuning dibandingkan	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
							para Kleting lainnya?	
Q36	E0	E0	E0	Accepted	The answer is fully consistent with information in several chunks and does not add interpretations outside the document.	No Discussion Needed	Nilai moral apa yang terkandung dalam penolakan Kleting Kuning terhadap permintaan Yuyu Kangkang?	Penolakan Kleting Kuning terhadap Yuyu Kangkang mencerminkan keberanian, integritas, dan keteguhan prinsip. Ia menolak kompromi moral, tidak menukar kehormatan dan harga diri demi kemudahan sesaat, serta memegang teguh kebenaran dan kejujuran.
Q37	E0	E0	E0	Accepted	The answer is consistent with the document content and directly supported by reference chunks without additional inference.	No Discussion Needed	Bagaimana karakter kesabaran Kleting Kuning digambarkan dalam cerita Ande-Ande Lumut?	Kleting Kuning digambarkan sebagai sosok yang sangat sabar. Ia menghadapi perlakuan kejam dari ibu dan saudara tirinya tanpa membalas dendam atau menyerah. Kesabaran dan ketabahannya dalam menanggung penderitaan, termasuk rintangan Yuyu Kangkang, pada akhirnya membuahkan hasil manis.
Q38	E0	E4	E1	Accepted	The core answer matches the document (oral inheritance), but there is an addition of modern adaptation not explicitly mentioned in the reference chunk.	No Discussion Needed	Bagaimana kisah Ande-Ande Lumut diwariskan dalam masyarakat Jawa menurut teks?	Kisah Ande-Ande Lumut diwariskan dalam masyarakat Jawa melalui penuturan lisan dan berbagai bentuk adaptasi, mulai dari tradisi lisan hingga seni pertunjukan dan media modern.

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
Q39	E2	E4	E1	Accepted	The core answer aligns with the document, but there is elaboration of additional moral values not explicitly mentioned in the reference chunk.	No Discussion Needed	Mengapa Ande-Ande Lumut dianggap penting dalam konteks budaya Nusantara?	Ande-Ande Lumut dianggap penting sebagai mahakarya budaya takbenda yang membentuk moral, etika, dan identitas bangsa melalui kekuatan narasi lisan.
Q40	E2	E4	E2	Accepted	The answer is thematically correct, but its use as modern school learning media is not directly supported by the reference chunk.	No Discussion Needed	Bagaimana kisah Ande-Ande Lumut dapat digunakan sebagai media pembelajaran karakter di sekolah modern?	Kisah Ande-Ande Lumut dapat digunakan untuk pembelajaran karakter karena mengandung nilai budi pekerti, kesabaran, ketabahan, keberanian, dan integritas yang dicontohkan melalui tokoh Kleting Kuning.
Q41	E0	E0	E2	Accepted	The answer is thematically correct (moral values appropriate), but linking it to globalization era challenges is not explicitly supported by the chunk.	No Discussion Needed	Apa relevansi cerita Ande-Ande Lumut terhadap tantangan moral generasi muda di era globalisasi?	Cerita Ande-Ande Lumut relevan karena mengajarkan prioritas budi pekerti di atas penampilan fisik, kesabaran dan ketabahan, serta keberanian dan integritas dalam menolak kompromi moral.
Q42	E2	E4	E2	Accepted	The answer is thematically accurate, but the concepts of national cultural resilience and modern adaptation are not explicitly supported by the chunk.	No Discussion Needed	Bagaimana peran cerita rakyat seperti Ande-Ande Lumut dalam memperkuat ketahanan	Cerita rakyat seperti Ande-Ande Lumut berperan sebagai media pewarisan nilai luhur dan pembentuk identitas budaya melalui kekuatan narasi lisan.

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
							budaya nasional?	
Q43	E2	E4 & E1	E2	Accepted	The answer is thematically accurate based on the character's patience values, but direct relevance to modern social life is not explicitly supported by the chunk.	No Discussion Needed	Mengapa nilai kesabaran yang ditunjukkan Kleting Kuning penting dalam kehidupan sosial masyarakat saat ini?	Nilai kesabaran Kleting Kuning penting karena menunjukkan ketekunan dan ketabahan dalam menghadapi perlakuan buruk dan rintangan tanpa membalas dendam hingga akhirnya membuahkan hasil positif.
Q44	E2	E0	E2	Accepted	The answer is thematically accurate; the existence of digital adaptations is supported, but the impact on cultural interpretation is inferential and not explicit in the chunk.	No Discussion Needed	Bagaimana adaptasi digital cerita Ande-Ande Lumut memengaruhi cara masyarakat memaknai budaya tradisional?	Adaptasi digital membantu memperluas jangkauan dan menjaga relevansi cerita di kalangan generasi yang akrab dengan teknologi.
Q45	E2	E4	E0	Accepted	The answer is consistent with the reference and does not exceed information stated in the document.	No Discussion Needed	Apa makna penting dari kisah Ande-Ande Lumut bagi masyarakat?	Kisah Ande-Ande Lumut memiliki makna penting sebagai cerminan nilai-nilai luhur budaya Jawa, berfungsi sebagai instrumen pendidikan karakter dan pewarisan nilai-nilai luhur antar generasi.
Q46	E2	E4 & E1	E1	Accepted	The core answer matches the document, but there is additional elaboration (bridge	No Discussion Needed	Bagaimana seharusnya kisah Ande-	Kisah Ande-Ande Lumut dipahami sebagai cerita bermuatan nilai luhur yang

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
					between generations and preservation of Nusantara identity) not explicitly stated in the chunk.		Ande Lumut dipahami oleh pembaca?	menekankan budi pekerti, kesabaran, ketabahan, serta keberanian dan integritas moral.
Q47	E2	E0	E0	Accepted	The answer is fully consistent with the reference and supported by several chunks without adding interpretations outside the document.	No Discussion Needed	Nilai apa saja yang bisa diambil dari tokoh Kleting Kuning?	Tokoh Kleting Kuning mengajarkan prioritas budi pekerti di atas penampilan fisik, kesabaran dan ketabahan dalam menghadapi cobaan, serta keberanian dan integritas dalam memegang prinsip moral.
Q48	E10	E10	E9	Rejected	There is no substantive answer, references, or supporting chunks. The system provides a fallback without content, so evaluation sets an abstain.	Discussion Needed	Mengapa cerita rakyat seperti Ande-Ande Lumut masih relevan hingga kini?	Tidak ada respon dari sistem saat ini, silahkan ajukan pertanyaan lainnya!
Q49	E0	E0	E0	Accepted	The answer is consistent with the reference and summarizes the main moral message without adding interpretations outside the document.	No Discussion Needed	Apa pesan utama yang ingin disampaikan melalui cerita Ande-Ande Lumut?	Pesan utama cerita Ande-Ande Lumut adalah pentingnya budi pekerti, kesabaran, ketabahan, serta keberanian dan integritas di atas penampilan fisik.
Q50	E0	E0	E9	Accepted	The document explicitly states the absence of specific historical information related to the kingdom period. The answer is an abstain and appropriate to the document's knowledge boundaries.	No Discussion Needed	Pada masa kerajaan apa kisah Ande-Ande Lumut pertama kali berkembang di Jawa?	Dokumen tidak menyebutkan secara spesifik masa kerajaan tempat kisah Ande-Ande Lumut pertama kali berkembang di Jawa.

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
Q51	E0	E0	E9	Accepted	The document does not provide information regarding the kingdom of origin of Ande-Ande Lumut. The answer states the absence of data and the system provides a fallback, so evaluation sets an abstain.	No Discussion Needed	Status sosial Ande-Ande Lumut sebagai pangeran berasal dari kerajaan mana menurut sejarah?	Dokumen yang diberikan tidak menyebutkan kerajaan asal Ande-Ande Lumut secara spesifik. Ia digambarkan sebagai seorang putra mahkota.
Q52	E0	E0	E0	Accepted	The answer is consistent and directly supported by the reference chunk without additional inference outside the document.	No Discussion Needed	Apakah kisah Ande-Ande Lumut memiliki latar waktu sejarah tertentu yang dapat diidentifikasi secara kronologis?	Kisah Ande-Ande Lumut tidak memiliki latar waktu sejarah tertentu yang dapat diidentifikasi secara kronologis dan digambarkan sebagai dongeng yang melampaui batasan waktu dan ruang.
Q53	E0	E0	E9	Accepted	The document does not provide information about the mythological origins of Yuyu Kangkang beyond narrative description. The answer correctly states the absence of data without speculation.	No Discussion Needed	Apakah tokoh Yuyu Kangkang terinspirasi dari makhluk mitologis tertentu dalam kepercayaan Jawa kuno?	Yuyu Kangkang digambarkan sebagai sosok mitologis berupa kepiting raksasa penjaga sungai, namun dokumen tidak menjelaskan inspirasi mitologis spesifik dalam kepercayaan Jawa kuno.
Q54	E0	E0	E9	Accepted	The document explicitly states the absence of information regarding official use as cultural propaganda. The	No Discussion Needed	Apakah cerita Ande-Ande Lumut pernah digunakan secara resmi	Tidak ada informasi dalam dokumen yang menyatakan penggunaan resmi cerita Ande-Ande Lumut sebagai

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
					answer appropriately refrains from speculation.		sebagai alat propaganda budaya oleh pemerintah?	alat propaganda budaya oleh pemerintah.
Q55	E0	E0	E9	Accepted	The document does not contain data about historians or written manuscripts; the answer correctly states the absence of information without speculation.	No Discussion Needed	Siapa tokoh sejarawan Jawa yang pertama kali mendokumentasikan kisah Ande-Ande Lumut dalam manuskrip tertulis?	Informasi mengenai tokoh sejarawan Jawa yang pertama kali mendokumentasikan kisah Ande-Ande Lumut dalam manuskrip tertulis tidak ditemukan dalam dokumen yang diberikan.
Q56	E0	E0	E9	Accepted	The document does not contain manuscript data, year, or location of discovery; the answer correctly states the absence of information without speculation.	No Discussion Needed	Sebutkan tahun dan lokasi ditemukannya naskah kuno Ande-Ande Lumut yang menjadi rujukan utama cerita ini.	Informasi mengenai tahun dan lokasi ditemukannya naskah kuno Ande-Ande Lumut tidak ditemukan dalam dokumen yang diberikan. Cerita ini lebih dikenal sebagai warisan budaya takbenda dan narasi lisan yang diwariskan secara turun-temurun.
Q57	E0	E0	E9	Accepted	The question requests historical genealogical data that is not available; abstain is appropriate to prevent hallucination of royal lineage.	No Discussion Needed	Bagaimana hubungan genealogis Ande-Ande Lumut dengan tokoh-tokoh kerajaan Jawa lainnya menurut	Dokumen yang diberikan tidak menjelaskan hubungan genealogis Ande-Ande Lumut dengan tokoh-tokoh kerajaan Jawa lainnya menurut catatan sejarah. Ande-Ande Lumut dijelaskan sebagai dongeng dan mahakarya takbenda.

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
							catatan sejarah?	
Q58	E0	E0	E9	Accepted	The question requests specific cultural rituals never mentioned in the document; abstain is necessary to prevent cultural inference without textual basis.	No Discussion Needed	Ritual budaya apa yang secara khusus terinspirasi dari kisah Yuyu Kangkang dalam tradisi Jawa kuno?	Berdasarkan dokumen yang diberikan, tidak ada informasi mengenai ritual budaya khusus yang terinspirasi dari kisah Yuyu Kangkang dalam tradisi Jawa kuno. Yuyu Kangkang disebutkan sebagai sosok mitologis dalam cerita Ande-Ande Lumut.
Q59	E0	E0	E9	Accepted	The question requests quantitative/textual classification (length of story versions) never explained in the document; abstain answer is appropriate and prevents speculation.	No Discussion Needed	Apakah dokumen ini menyebutkan versi terpendek atau terpanjang dari cerita Ande-Ande Lumut?	Dokumen ini tidak menyebutkan versi terpendek atau terpanjang dari cerita Ande-Ande Lumut. Dokumen hanya membahas esensi cerita dan adaptasinya dalam berbagai bentuk.
Q60	E7	E7	E8	Rejected	The question contains anachronism (modern technology element) irrelevant to the traditional story context; the model successfully corrects the assumption and returns to the narrative context.	Discussion Needed	Mengapa para Kleting menyeberang sungai tidak menyewa speedboat?	Para Kleting tidak menyeberang sungai menggunakan speedboat karena cerita Ande-Ande Lumut adalah dongeng tradisional yang berlatar waktu sebelum teknologi modern seperti speedboat ada. Penyeberangan sungai dalam cerita tersebut melibatkan sosok mitologis Yuyu Kangkang.

Query_ID	Evaluator_1_Score	Evaluator_2_Score	Evaluator_3_Score	Final_Decision	Annotation_Notes	Resolution_Type	User_Query (Indonesian)	Candidate_Response (Indonesian)
Q61	E4	E4	E2	Rejected	The answer correctly states the absence of information, but adds context about Yuyu Kangkang not requested in the question, thus going beyond the query scope.	Discussion Needed	Siapa yang menjadi pelanggan rutin menyeberang dan apakah mendapatkan potongan tarif?	Berdasarkan dokumen yang diberikan, tidak ada informasi mengenai siapa yang menjadi pelanggan rutin penyeberangan atau apakah mereka mendapatkan potongan tarif. Penyeberangan sungai yang disebutkan hanya melibatkan Yuyu Kangkang yang menawarkan jasa dengan imbalan ciuman.