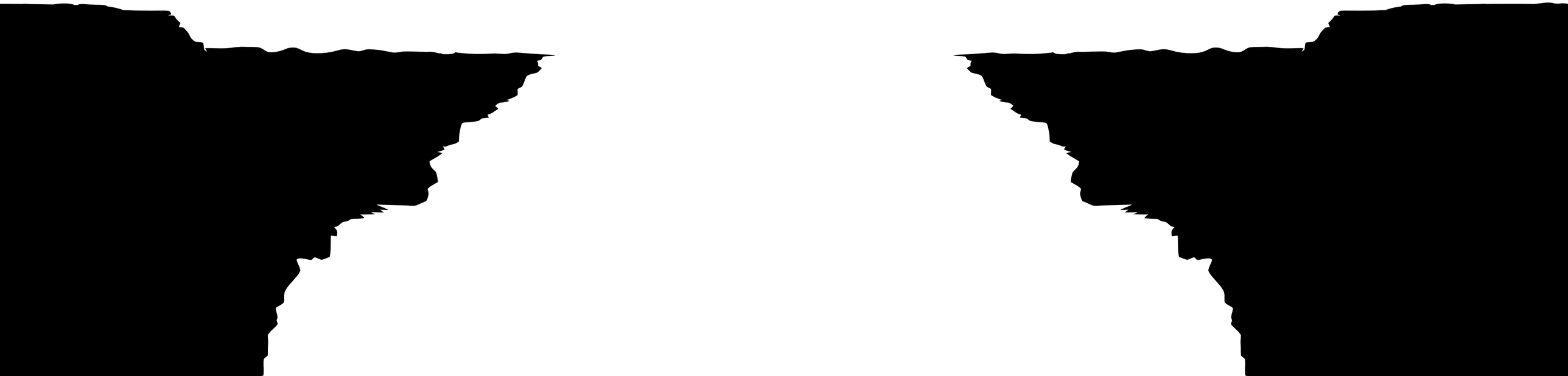
STARK-QA

(Story Text And Related Knowledge Question Answering)

Team Name: HansRaj

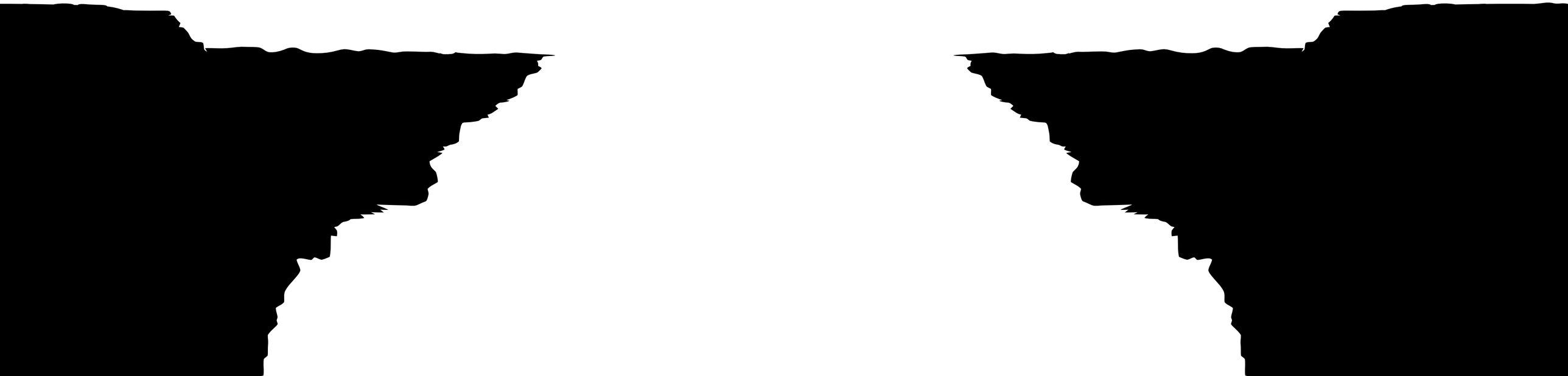
Archit Choudhary
Ishaan Romil
Sujal Deoda

LITERATURE REVIEW



LITERATURE REVIEW

RAG

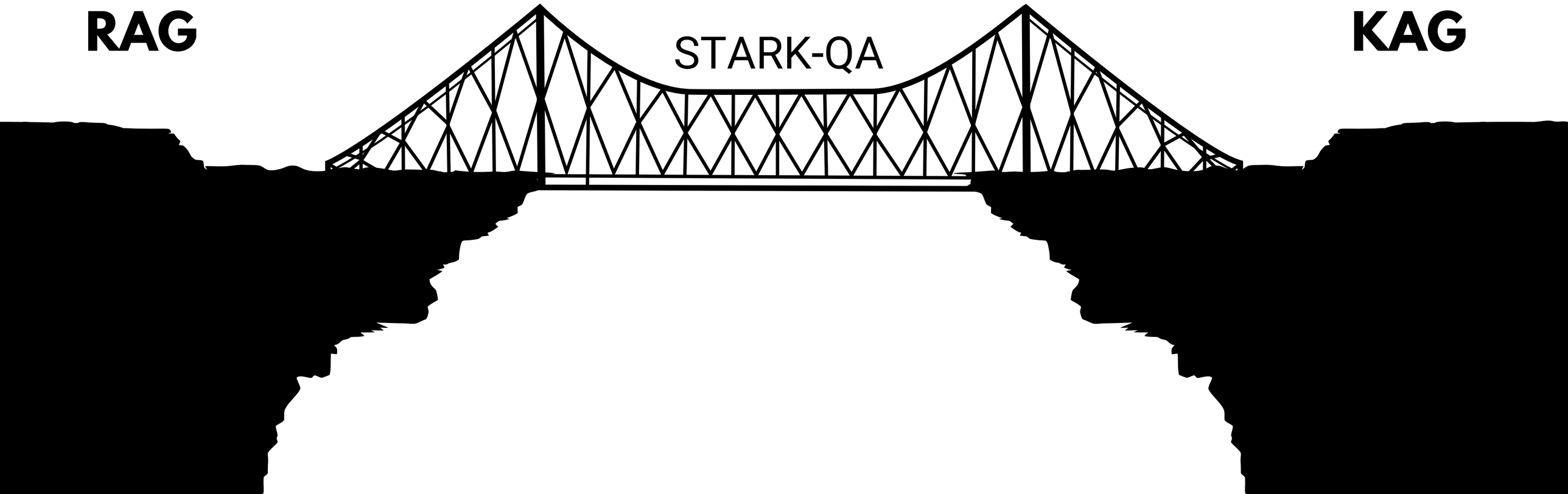


LITERATURE REVIEW

RAG

KAG

LITERATURE REVIEW



WHAT WE PROPOSE?

PROJECT OUTLINE

UNSTRUCTURED

TEXTUAL

WHAT WE PROPOSE?

PROJECT OUTLINE

UNSTRUCTURED

TEXTUAL

STRUCTURED

GRAPHICAL

WHAT WE PROPOSE?

PROJECT OUTLINE

A dual-pipeline question answering system designed specifically for literary works that combines the power of Retrieval Augmented Generation (RAG) and Knowledge Augmented Generation (KAG) techniques.

UNSTRUCTURED

TEXTUAL

STRUCTURED

GRAPHICAL

WHAT WE PROPOSE?

PROJECT OUTLINE

This system is specifically designed for processing and analyzing short stories (under 2,500 words) using knowledge graphs and retrieval augmented generation. The implementation consists of two main components that work together to understand narrative structures and enhance AI interactions with literary content.

UNSTRUCTURED

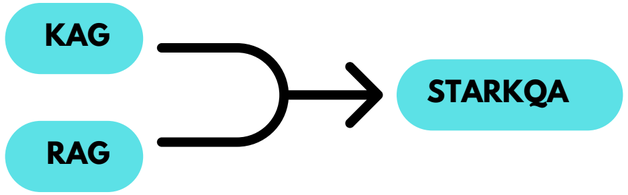
TEXTUAL

STRUCTURED

GRAPHICAL

WHAT WE PROPOSE?

ARCHITECTURE



Initial Hypothesis, combining KAG and RAG, would improve the quality of responses that we get.

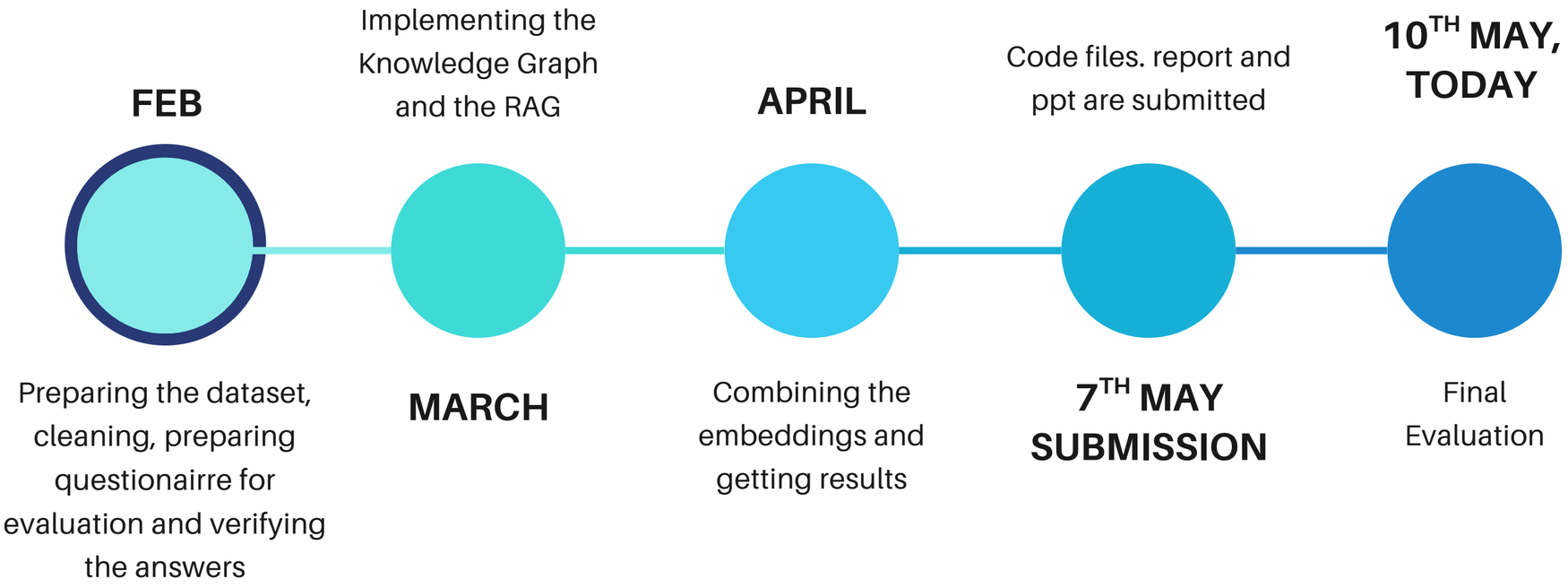
On implementation with the baseline models only a very minimal improvement was seen.

Thus, we went about and made certain optimisations.

On making the optimisations. the combined system now gave some significantly improved responses.

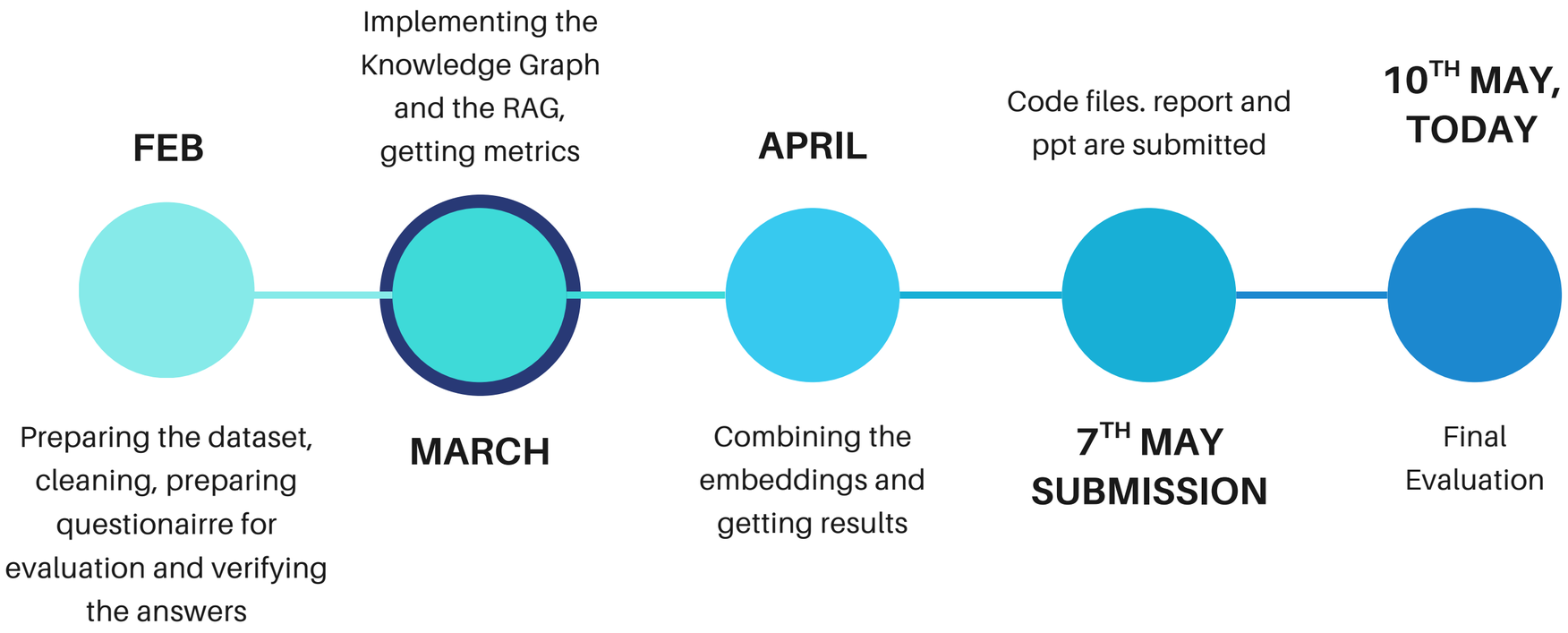
DID WE DO THINGS AS PLANNED?

TIMELINE



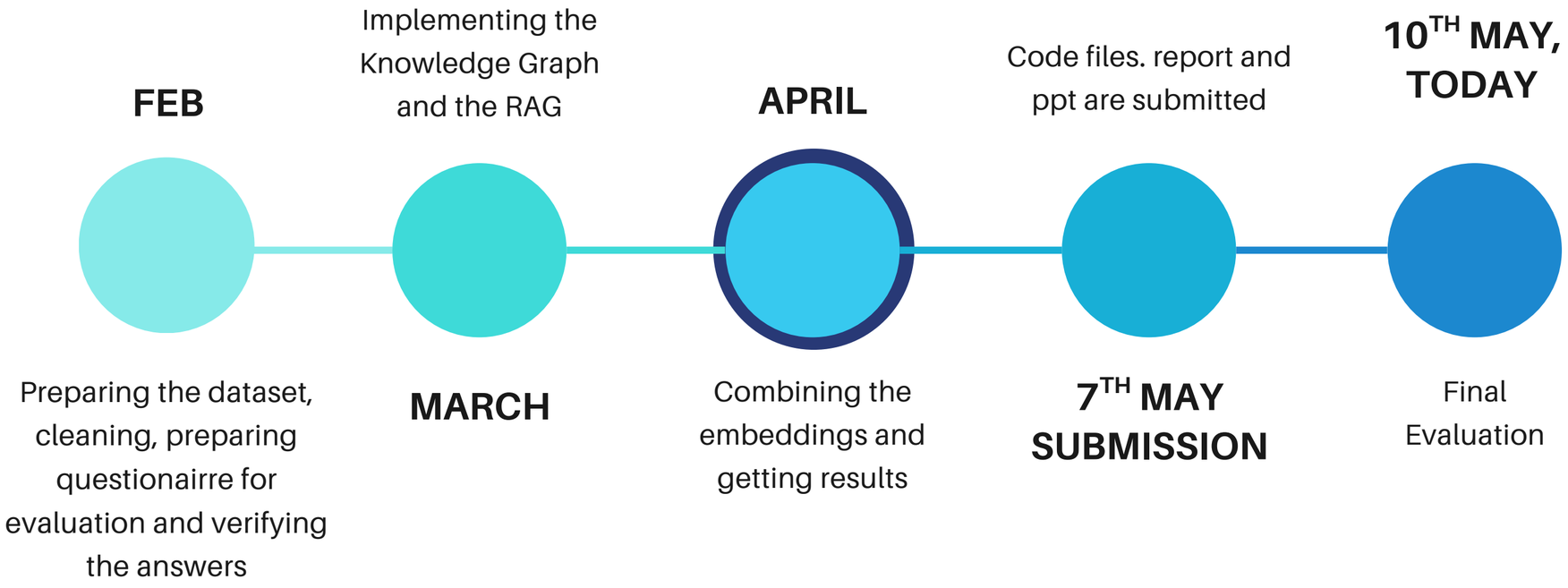
DID WE DO THINGS AS PLANNED?

TIMELINE



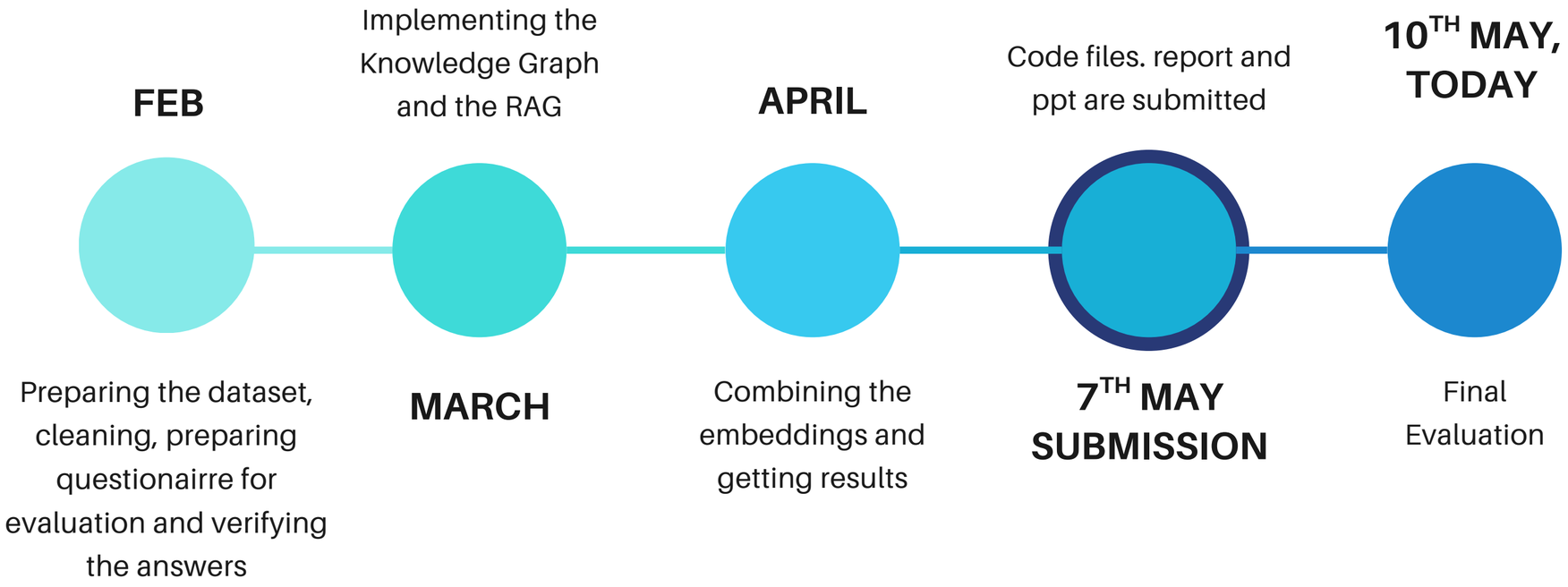
DID WE DO THINGS AS PLANNED?

TIMELINE



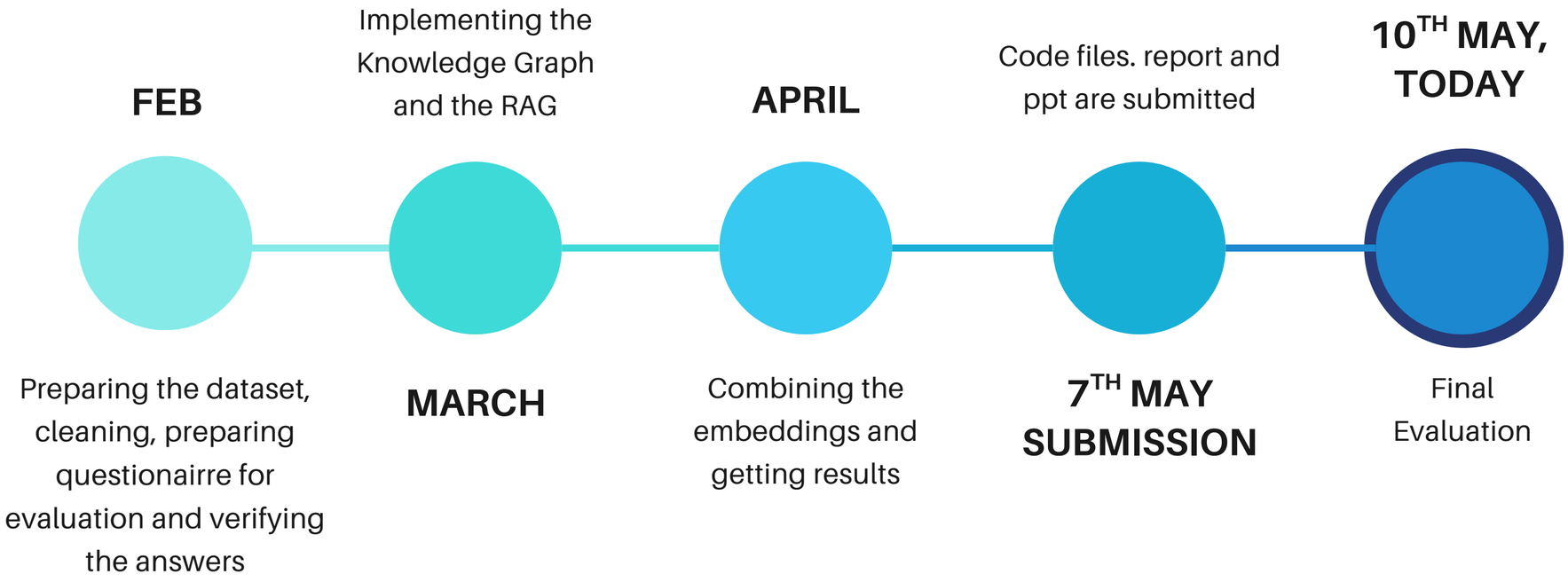
DID WE DO THINGS AS PLANNED?

TIMELINE



DID WE DO THINGS AS PLANNED?

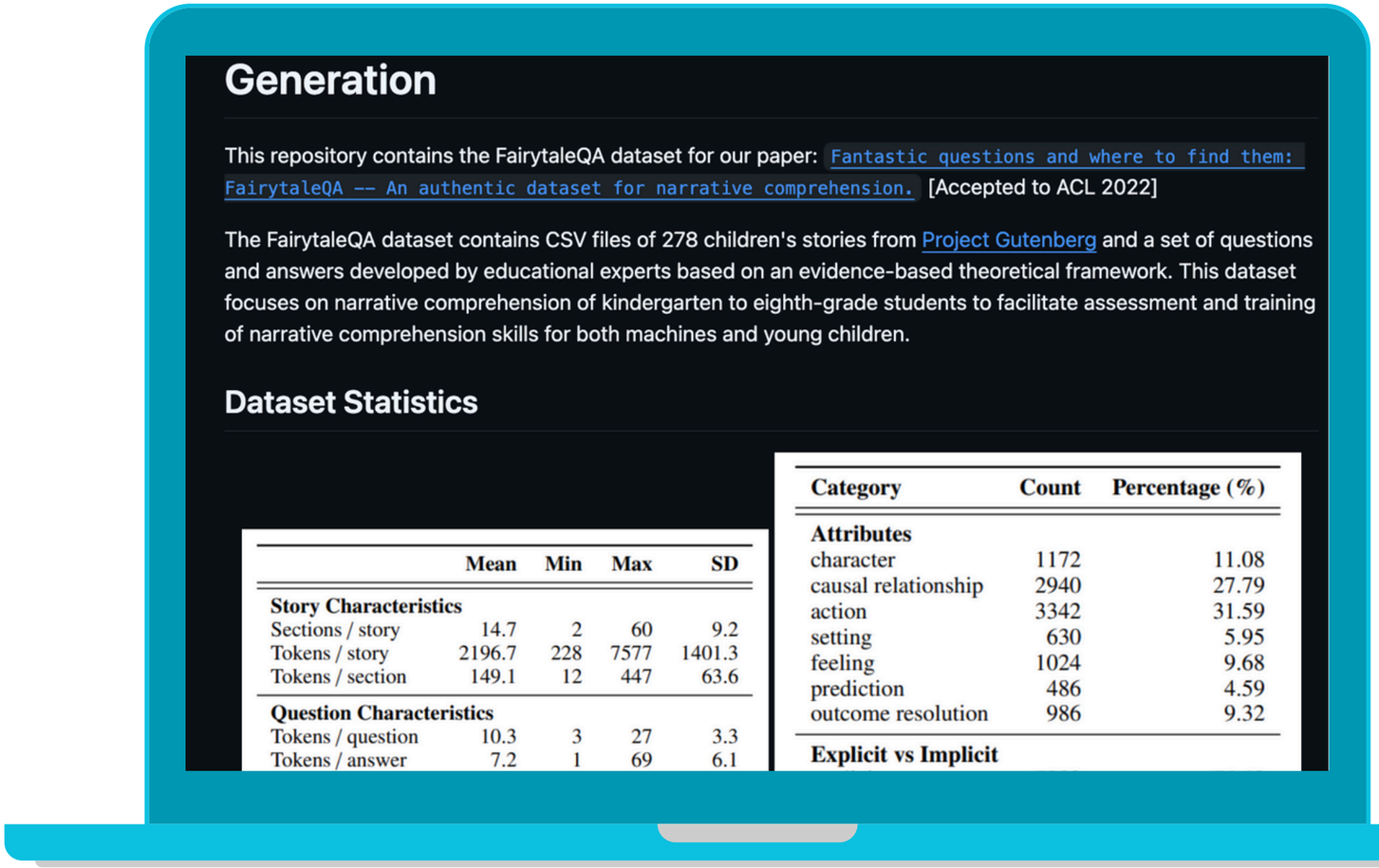
TIMELINE



ABOUT OUR
DATASET

The FairytaleQA dataset contains CSV files of 278 children's stories from Project Gutenberg and a set of questions and answers developed by educational experts based on an evidence-based theoretical framework. This dataset focuses on narrative comprehension of kindergarten to eighth-grade students to facilitate assessment and training of narrative comprehension skills for both machines and young children.

We tested on 100 questions from the set of 10000 due to time constraints and compute issues.



https://github.com/uci-soe/FairytaleQAData

ABOUT OUR
RAG

Retrival Augmented Generation

Baseline:

- Basic document chunking (300 words)
- vectorisation using tf-idf
- cosine similarity for retrieval top-k for most similar chunks
- and simple intersection with questions for generating responses

question_id	question	ground_truth	prediction	question_type	local_or_sum	ex_or_im	prediction_time
0	where did the old man and his wife live ?	in a dear little cottage by the side of a burn .	there was once an old man and his wife , who lived in a dear little cottage by the side of a burn	setting	local	explicit	0.0019896030
1	who were a very canty and contented couple ?	an old man and his wife .	they were a very canty and contented couple , for they had enough to live on , and enough to do	character	local	explicit	0.00218057632
2	what animals did the couple own ?	two sleek cows , five hens and a cock , an old cat , and two	' ' how did you get the ring which was in the soup ? ' he asked	action	local	explicit	0.0035600662
3	what did the old man spend his time doing ?	looking after the cows , and the hens , and the garden .	the old man spent his time looking after the cows , and the hens , and the garden ; while the old woman kept herself busy spinning	action	local	explicit	0.00402784347
4	what did the old woman want for supper ?	an oatmeal bannock .	one day , just after breakfast , the old woman thought that she would like an oatmeal bannock for her supper that evening , so she took down her bakeboard , and put on her girdle , and baked a couple of fine cakes , and when they were ready she put them down before the fire to harden	action	local	explicit	0.00173020362

Fig. Predictions of Basic RAG

ABOUT OUR **RAG**

Retrival Augmented Generation

Optimisation:

- combining two techniques for retrieval both sparse and dense
- Fusional strategy
- overlapping chunking, semantic contInuity
- question types identification.
- reranking based on question type specific boosting

question_id	question	ground_truth	prediction	question_type	local_or_sum	ex_or_im	prediction_time
0	where did the old man and his wife live ?	in a dear little cottage by the side of a burn .	there was once an old man and his wife , who lived in a dear little cottage by the side of a burn	setting	local	explicit	0.0019896030
1	who were a very canty and contented couple ?	an old man and his wife .	they were a very canty and contented couple , for they had enough to live on , and enough to do	character	local	explicit	0.00218057632
2	what animals did the couple own ?	two sleek cows , five hens and a cock , an old cat , and two	' ' how did you get the ring which was in the soup ? ' he asked	action	local	explicit	0.0035600662
3	what did the old man spend his time doing ?	looking after the cows , and the hens , and the garden .	the old man spent his time looking after the cows , and the hens , and the garden ; while the old woman kept herself busy spinning	action	local	explicit	0.00402784347
4	what did the old woman want for supper ?	an oatmeal bannock .	one day , just after breakfast , the old woman thought that she would like an oatmeal bannock for her supper that evening , so she took down her bakeboard , and put on her girdle , and baked a couple of fine cakes , and when they were ready she put them down before the fire to harden	action	local	explicit	0.00173020362

Fig. Predictions of Basic RAG

BASIC KAG IMPLEMENTATION

- **Entity Extraction:** Identifies characters, places, and objects from story texts using both spaCy NER and regex patterns tailored for fairytales.
- **Triple Extraction:** Parses each sentence to extract subject–predicate–object relationships using dependency rules (e.g., king fought dragon, queen's crown → queen has crown).
- **Graph Construction:** Builds a directed knowledge graph where entities are nodes and edges represent typed relationships (e.g., has, located_in), with story metadata attached.
- **Sentence Mapping:** Links each extracted triple to its original sentence and story, enabling traceability and direct evidence retrieval during question answering.
- **Question Answering:** Detects relevant entities in a user question, finds matching triples, and generates contextual answers by retrieving the most relevant supporting sentence from the graph.

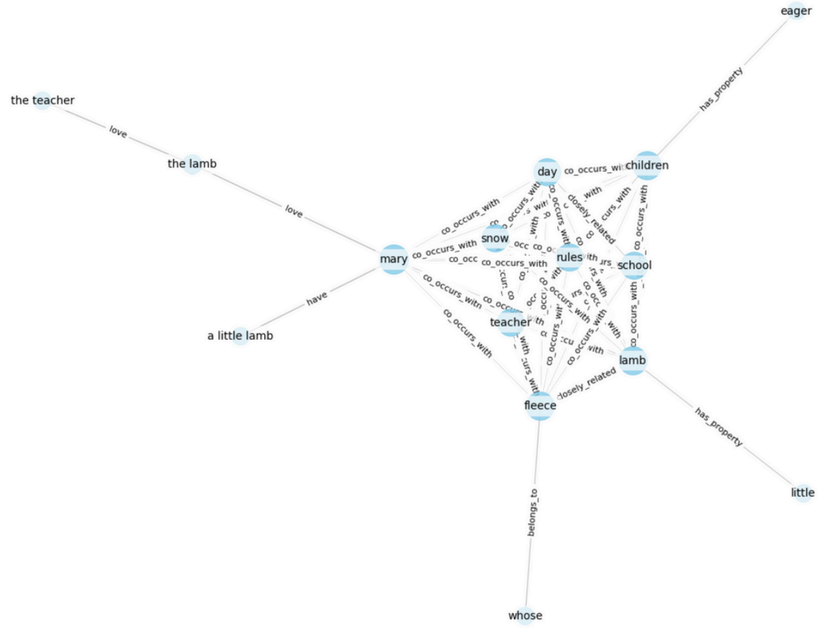


Fig. Knowledge graph generated for the Mary and the Lamb Story

IMPROVING
KAG

- **Expanded Entity Coverage:** Adds detailed object categories, animals, emotions, household items, and adjective–noun phrases (e.g., golden hair, angry king)
- **Entity Type Tagging and Frequency Tracking:** Annotates each entity with its type (PERSON, OBJECT, CONCEPT) and counts its frequency in the story
- **Flexible Semantic Matching for Questions:** Supports semantic fallbacks (e.g., wife → queen, woman) and ranks context sentences by overlap with question words
- **Triple-Context Fusion:** Combines extracted relationships with full sentence-level context to improve answer richness, especially when multiple entities or story segments are involved.

question_id	question	ground_truth	prediction	question_type	local_or_sum	ex_or_im	prediction_tim
0	where did the old man and his wife live ?	in a dear little cottage by the side of a burn .	This takes place in various locations described in the fairytales.	setting	local	explicit	0.014204740
1	who were a very canty and contented couple ?	an old man and his wife .	This question relates to husband, which appears in: the-wee-bannock, alleleiraugh-or-the-many-	character	local	explicit	0.009068250
2	what animals did the couple own ?	two sleek cows , five hens and a cock , an old cat , and two kittens .	This question relates to father, which appears in: alleleiraugh-or-the-many-furred-	action	local	explicit	0.007905483
3	what did the old man spend his time doing ?	looking after the cows , and the hens , and the garden .	This relates to objects and events described in the fairytales.	action	local	explicit	0.009130954
4	what did the old woman want for supper ?	an oatmeal bannock .	This relates to objects and events described in the fairytales.	action	local	explicit	0.008457660

A STEP AHEAD

STARK-QA

The system introduces a hybrid approach that combines Knowledge Augmented Generation (KAG) and Retrieval Augmented Generation (RAG) to enhance question answering. It processes stories in parallel—extracting structured knowledge graphs and chunking raw text—and fuses both sources during query response. When asked a question, it retrieves key relationships from the graph and relevant text segments, then merges them into a coherent prompt. This dual-retrieval strategy boosts accuracy and depth by uniting structured precision with unstructured nuance for more context-aware responses.

question_id	question	ground_truth	prediction	question_type	local_or_sum	ex_or_im	prediction_time
0	where did the old man and his wife live ?	in a dear little cottage by the side of a burn .	In a dear little cottage by.	setting	local	explicit	2.970161437988
1	who were a very canty and contented couple ?	an old man and his wife .	An old man and his wife.	character	local	explicit	1.638899803161
2	what animals did the couple own ?	two sleek cows , five hens and a cock , an old cat , and two kittens .	Two sleek cows , five hens.	action	local	explicit	1.702573060989
3	what did the old man spend his time doing ?	looking after the cows , and the hens , and the garden .	Two sleek cows , five hens.	action	local	explicit	1.858664512634
4	what did the old woman want for supper ?	an oatmeal bannock .	An oatmeal bannock.	action	local	explicit	1.335773944854

Fig. Predictions of STARK-QA

RESULTS AND

EVALUATION METRICS

We evaluated the KAG+RAG hybrid model on these criterias:

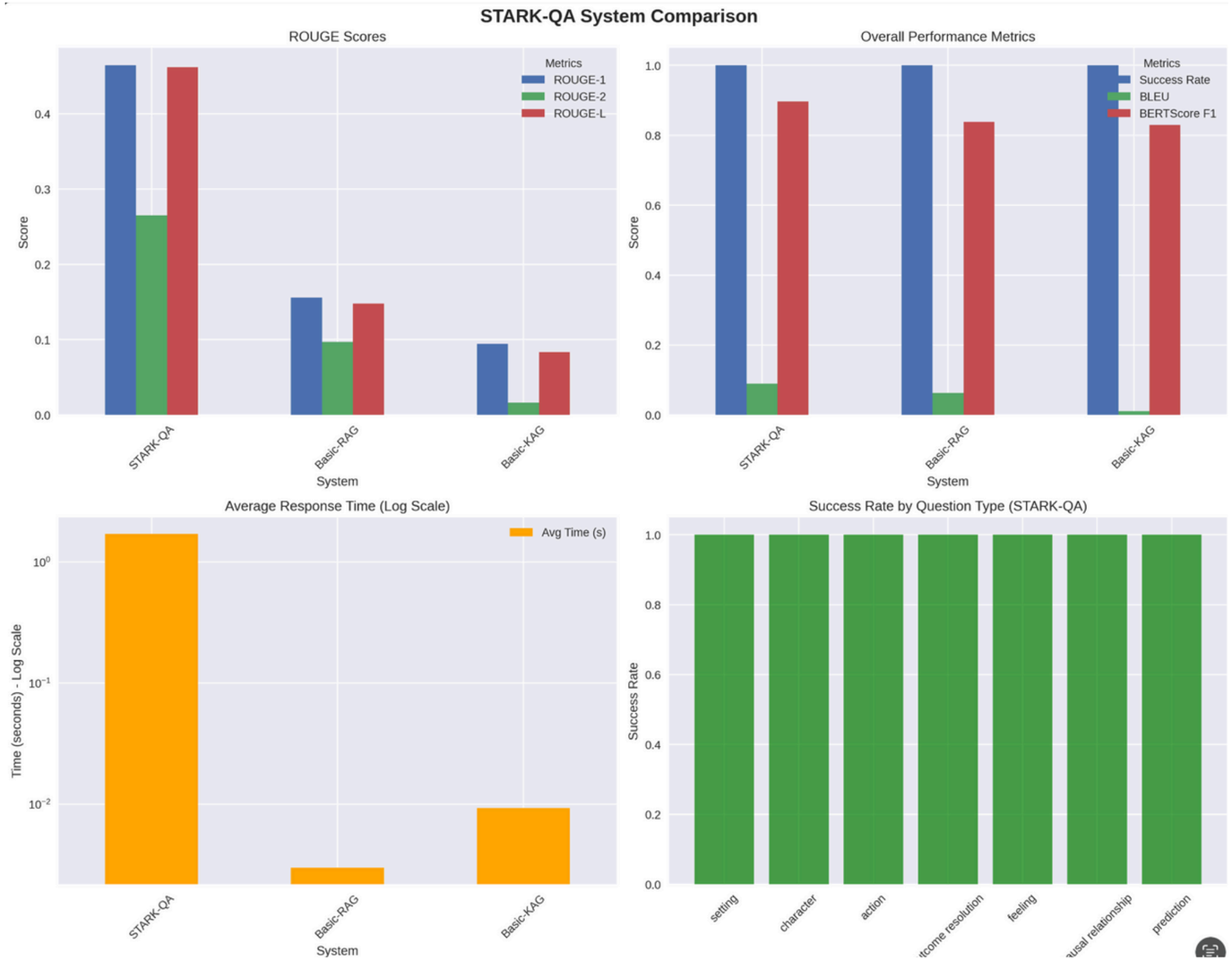
- Rouge
- BLEU
- BERTScore F1
- Average response time

This allows for nuanced assessment of story-grounded reasoning, valid paraphrasing, and semantic correctness, as illustrated in our sample evaluation template comparing outputs from the basic KAG, basic RAG, and their integration.

RESULTS

system_comparison

System	Total Questions	Valid Predictions	Success Rate	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	BERTScore F1	Avg Time (s)	Total Time (s)
STARK-QA	100	100	1.0	0.464365337484957 0	0.2649767739999940	0.46188413447743800	0.08969059823492190	0.896839439868927	1.7028361654281600	170.28361654281600
Basic-RAG	100	100	1.0	0.15575817925916200	0.09675072658614360	0.14767160220136400	0.06230673572434080	0.8385589122772220	0.002999880313873290	0.2999880313873290
Basic-KAG	100	100	1.0	0.09446393107968560	0.016178722847201100	0.08314111274953560	0.010756240839538800	0.829843282699585	0.009320721626281740	0.9320721626281740



CHALLENGES

1

Model Capacity Constraints: Mistral, while accessible, lacks the advanced reasoning and narrative understanding of larger models like GPT-4, limiting performance on complex questions.

2

Knowledge Graph Extraction Errors: The KAG module relies on syntactic parsing and NER, which can misidentify entities or relationships, leading to incomplete or incorrect graphs, thus we had to tailor the knowledge graph dependent on our dataset.

3

R**etrieval Quality Issues:** FAISS-based retrieval in the RAG module may return passages that are topically related but not contextually relevant, especially in stories with motif repetition or ambiguous characters.

4

Evaluation Challenges: We were initially using BLEU as an evaluation metric, but our results were inconsistent and contradictory. However, we realized it gives too much relevance to exact matches, which are difficult to achieve from non-deterministic LLMS.

THANK YOU