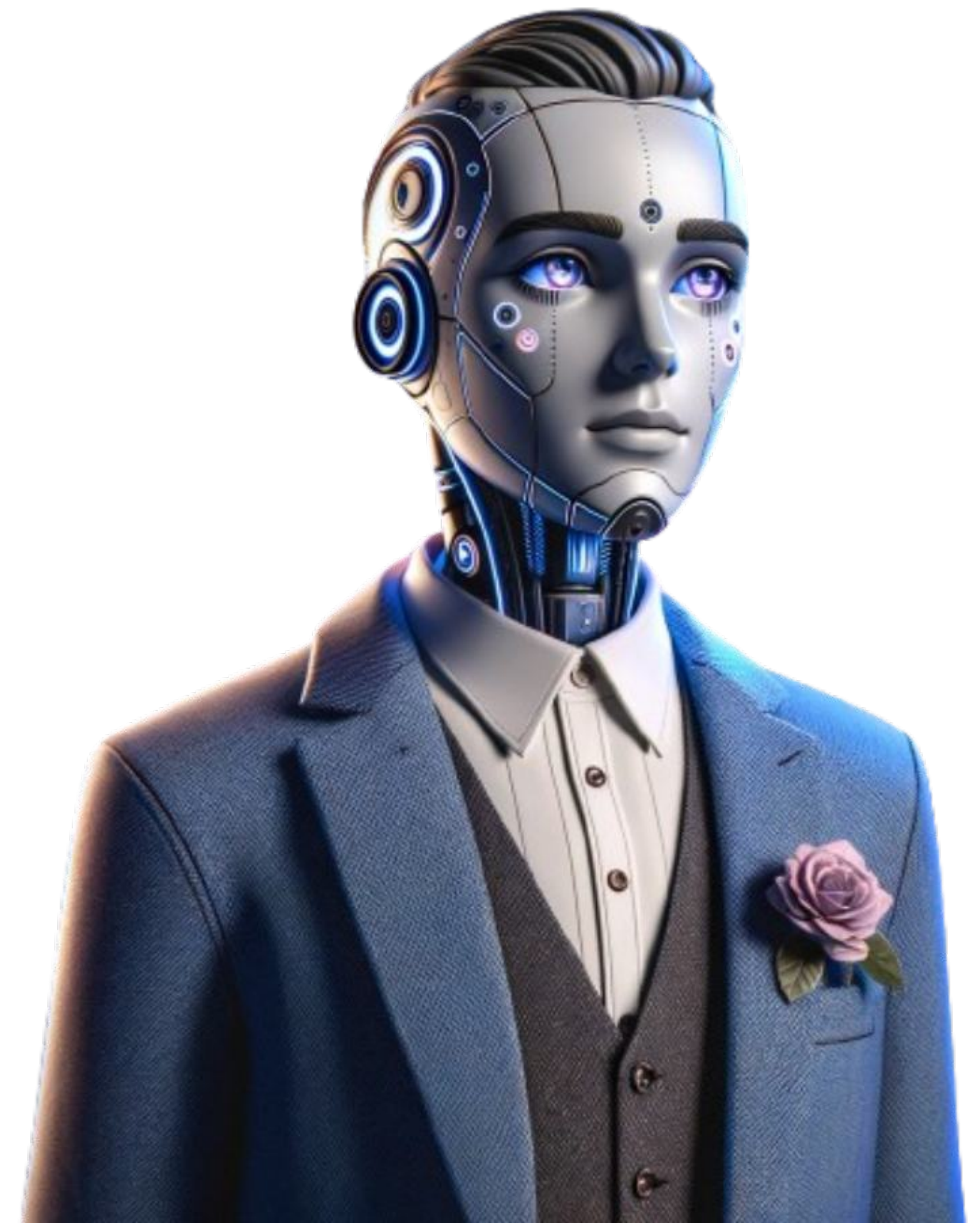


Kedward

Sistema RAG para responder perguntas sobre
o livro Cosmos, de Carl Sagan



Sobre o desafio...

Objetivo

Criar um sistema de perguntas e respostas baseado no livro Cosmos, utilizando a metodologia RAG (Retrieval-Augmented Generation).

Regras

- A avaliação será realizada utilizando o framework RAGAS *
- Os componentes e técnicas são de livre escolha, desde que incluam:
 - *Chat Input*
 - *Chat Output*
 - *Langwatch Evaluator*

A competição

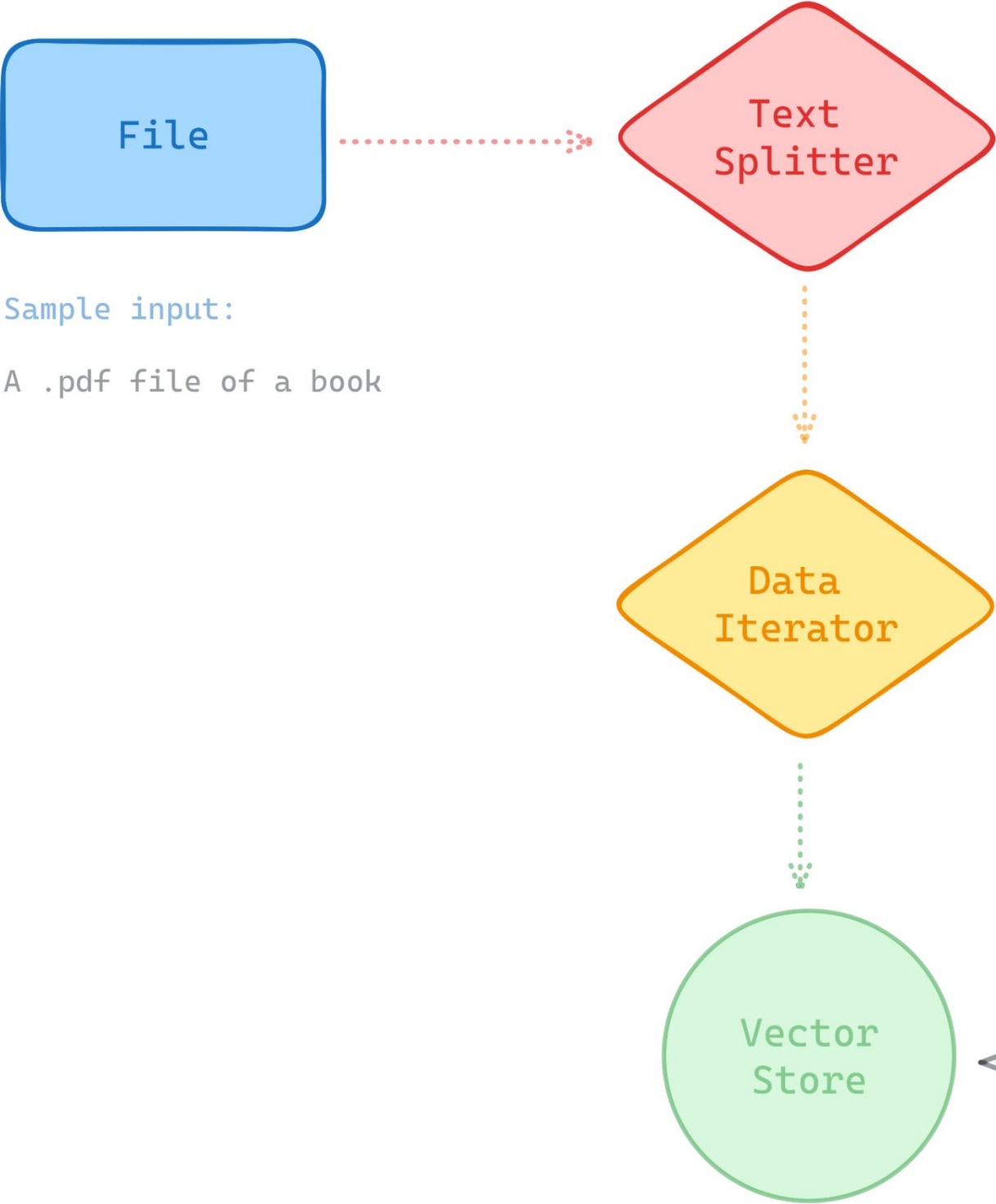
IADevs do Langflow é uma competição promovida pelo Langflow, focada em Inteligência Artificial, que fomenta talentos e fortalece a comunidade open-source no Brasil.

* Retrieval-Augmented Generation Assessment

Arquitetura

Cosmos RAG, Ingestion pipeline

by Naomi Lago, 2024



Sample input:
A .pdf file of a book

NOTES:

File:Currently, only supports .pdf

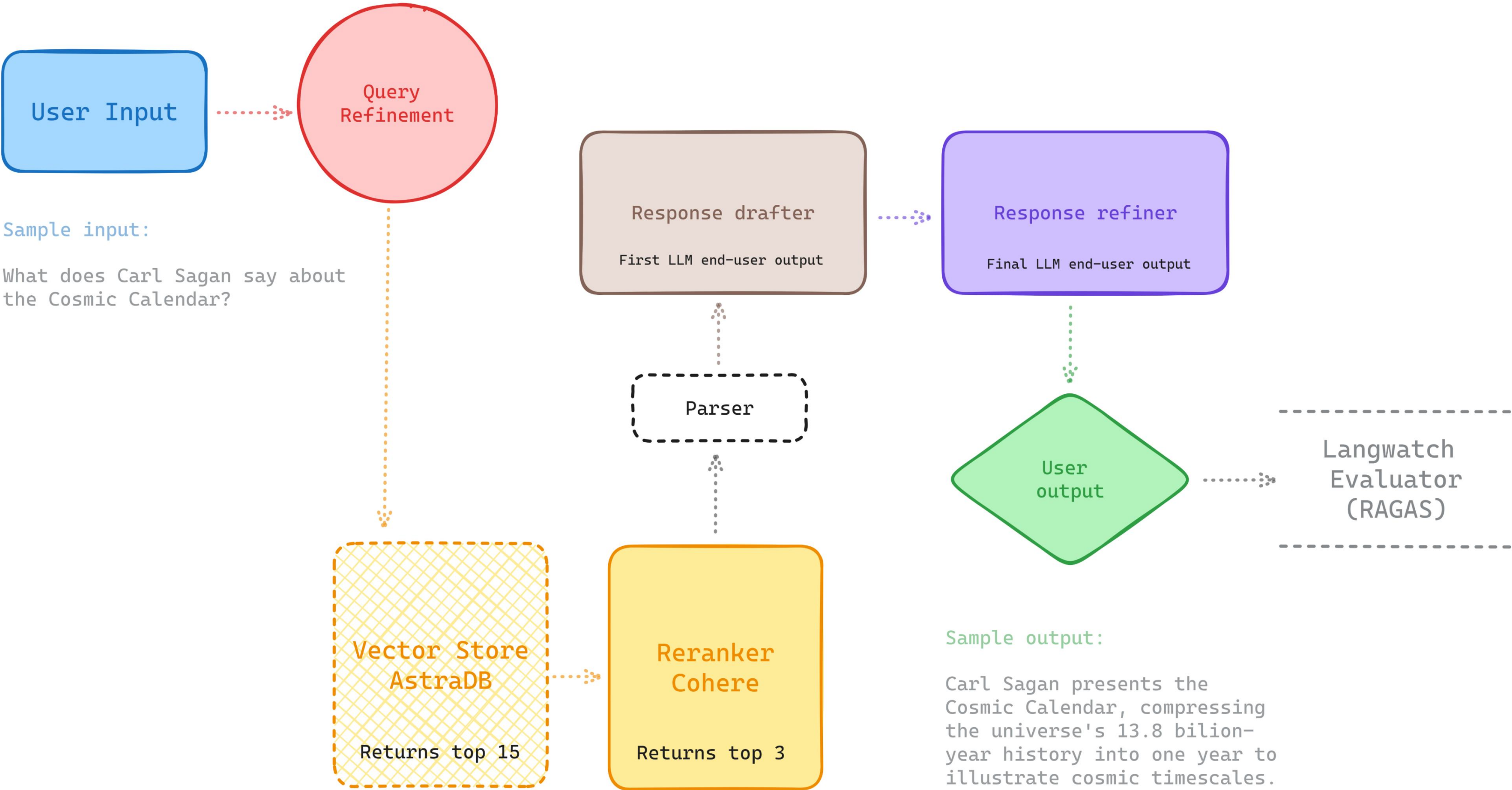
Text Splitter: Chunk size of 1000 and Overlap of 100

Data Iterator: Adding LLM generated metadata fields to chunk text's

Vector Store: Sending the preprocessed chunks to a vector database

Cosmos RAG, Inference pipeline

by Naomi Lago, 2024



Tecnologias



- **GPT-4o Mini:** Gerar os metadados durante a ingestão
- **Text Embedding 3 Large:** Gerar os embeddings para o banco vetorial

ANTHROPIC

- **Claude 3.5 Sonnet:** Refinar perguntas dos usuários, criar rascunhos de respostas e gerar o output final



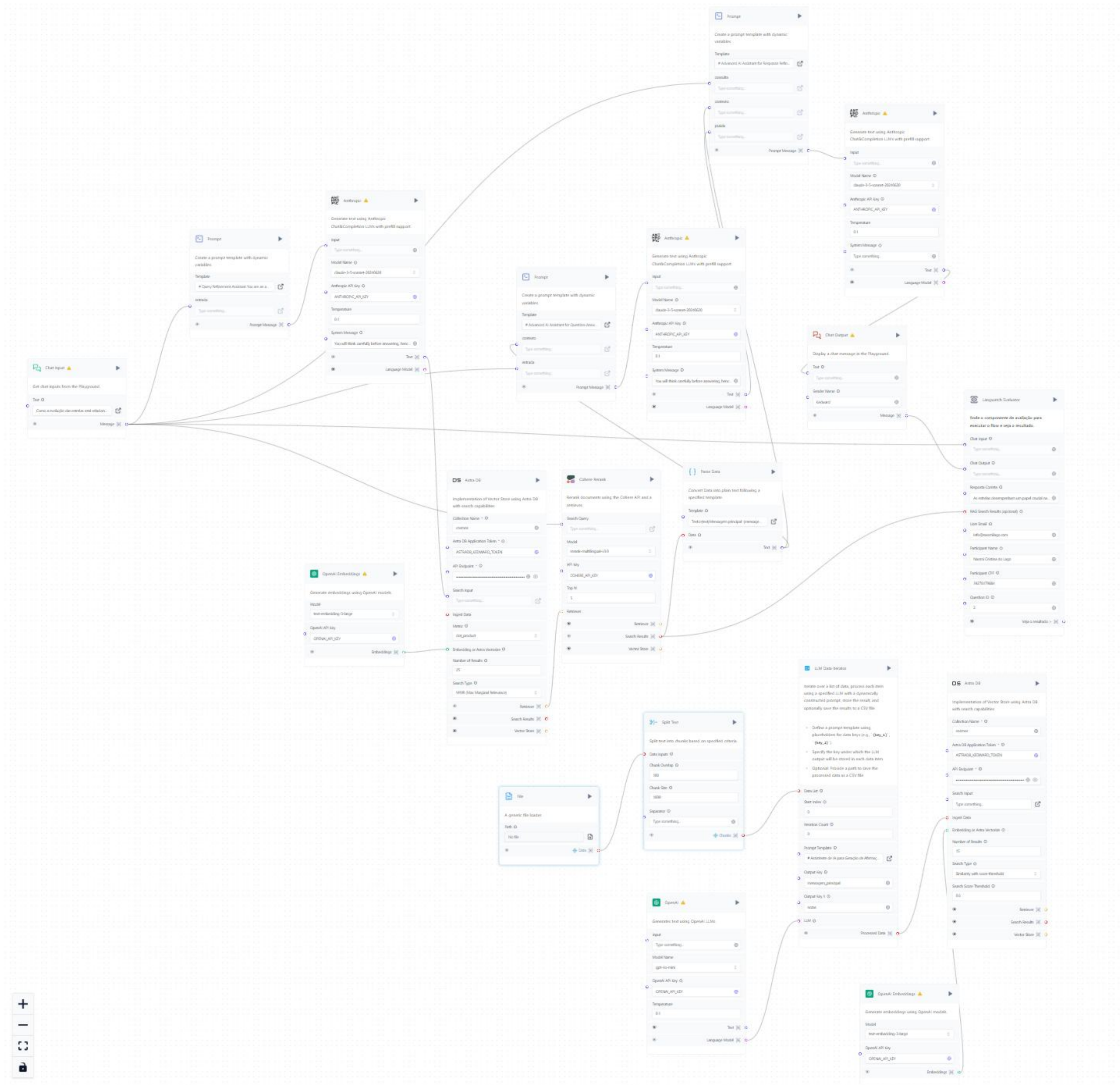
- **AstraDB:** Armazenar e pesquisar vetores de contexto e seus respectivos metadados



- **Rerank Multilingual v3.0:** Promover ranqueamento de documentos, ordenando por relevância



- Monitoramento e avaliação de perguntas e respostas



Organização

Google Sheets

Um grande aliado no monitoramento das métricas.

id	question	ground_truth
1	quantos cortes são necessários para dividir uma torta de maçã até chegar a um átomo?	Cerca de noventa cortes.
2	o que mudou na percepção de Júpiter após o encontro da Voyager 1, segundo o autor?	Júpiter passou a ser visto como um lugar a ser explorado, não apenas um ponto de luz.
3	qual é o principal risco de fingir irracionalidade para um blefe nuclear eficaz?	Acostumar-se com a irracionalidade e deixar de ser fingimento.
4	segundo Farrington, qual fator econômico levou ao declínio da ciência jônica?	A economia escravista.
5	qual foi o problema enfrentado pela Voyager 1 que poderia ter comprometido a coleta de dados científicos?	Problema no posicionamento da lança da plataforma de varredura científica.
6	o que Kepler acreditava sobre a geometria em relação à criação do universo?	Kepler acreditava que a geometria existiu antes da Criação e era coeterna com a Mente de Deus.
7	qual foi o impacto da rejeição da existência do éter luminífero" na compreensão da propagação da luz?"	Mostrou que a luz viaja no vácuo, fundamentando a teoria da relatividade de Einstein.
8	Qual é a analogia usada no texto para descrever a percepção de objetos ao viajar próximo à velocidade da luz?	Os objetos aparecem comprimidos numa pequena janela circular à sua frente.
9	Qual é a opinião de Carl Sagan sobre a possibilidade de formas de vida baseadas em elementos diferentes do carbono e água?	Sagan acha a possibilidade interessante, mas prefere carbono e água pela abundância e adequação para a vida.
10	qual é a visão perturbadora e instigante sobre a hierarquia de universos mencionada no texto?	A visão é de uma hierarquia infinita de universos, onde cada partícula pode ser um universo, e nosso universo pode ser uma partícula em um universo maior.

Q&A	AC 1 (baseline)	AC 2	AVG	DIFF	% CHANGE	FEEDBACK
1	0.88	0.91	0.90	0.03	3.41	Minor
2	0.55	0.70	0.63	0.15	27.27	Moderate
3	0.21	0.92	0.57	0.71	338.10	Exceptional
4	0.96	0.34	0.65	-0.62	-64.58	Decline
5	0.96	0.85	0.91	-0.11	-11.46	Decline
6	0.22	0.32	0.27	0.10	45.45	Moderate
7	0.91	0.92	0.92	0.01	1.10	Minor
8	0.51	0.91	0.71	0.40	78.43	Major
9	0.91	0.94	0.93	0.03	3.30	Minor
10	0.91	0.95	0.93	0.04	4.40	Minor
SUM	7.02	7.76	7.39	0.74	425.41	
AVERAGES	0.70	0.78	0.74	0.07	42.54	

Q&A	AC 2 (baseline)	AC 3	AVG	DIFF	% CHANGE	FEEDBACK
1	0.91	0.95	0.93	0.04	4.40	Minor
2	0.70	0.35	0.53	-0.35	-50.00	Decline
3	0.92	0.90	0.91	-0.02	-2.17	Decline
4	0.34	0.96	0.65	0.62	182.35	Exceptional
5	0.85	0.96	0.91	0.11	12.94	Moderate
6	0.32	0.37	0.35	0.05	15.63	Moderate
7	0.92	0.94	0.93	0.02	2.17	Minor
8	0.91	0.91	0.91	0.00	0.00	Minor
9	0.94	0.94	0.94	0.00	0.00	Minor
10	0.95	0.94	0.95	-0.01	-1.05	Decline
SUM	7.76	8.22	7.99	0.46	164.26	
AVERAGES	0.78	0.82	0.80	0.05	16.43	

Muito obrigada!

LinkedIn: <https://linkedin.com/in/naomilago>

E-mail: info@naomilago.com

Discord: @naomilake