UNIFI VALUE FRAMEWORKS PDF LIFTING COMPETITION:

TOP 2 SOLUTION

Unifi Value Frameworks PDF Lifting Competition

Can you lift existing ESG metrics from static PDF documents?

Intermediate   Generative AI   Sustainability

LlamaParse

(Recently released : Feb 20, 2024)

LlamaParse is a state-of-the-art parser designed to specifically unlock RAG over complex PDFs with embedded tables and charts.

Metrics	Baseline (PyPDF + Naive RAG)	LlamaParse + Recursive Retrieval
mean_correctness_score	3.874	4.270
mean_relevancy_score	0.844	0.926
mean_faithfulness_score	0.667	0.915

This service is available in a public preview mode: available to everyone, but with a usage limit (1k pages per day).

3$/1k pages
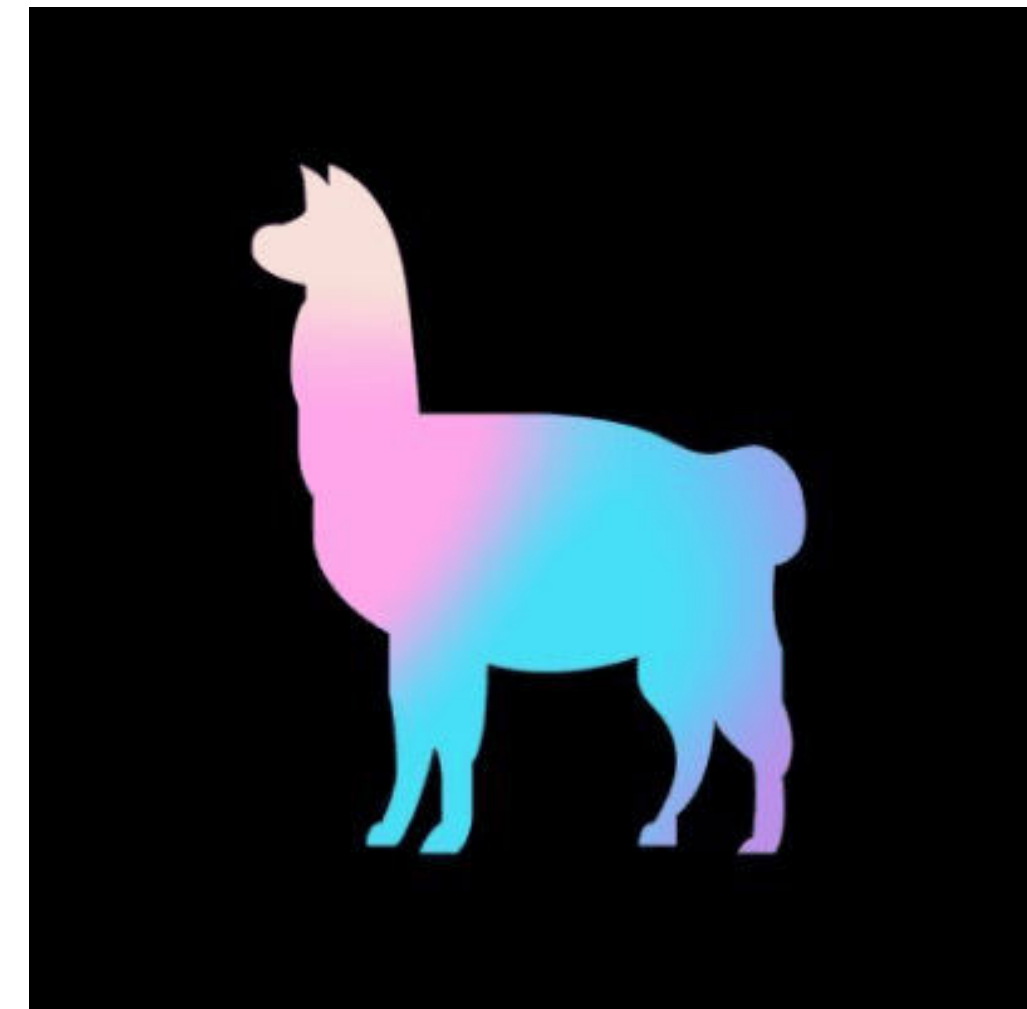
Hugging Face

MODEL : bart-large-cnn

BART is a transformer encoder-encoder (seq2seq) model with a bidirectional (BERT-like) encoder and an autoregressive (GPT-like) decoder. BART is pre-trained by (1) corrupting text with an arbitrary noising function, and (2) learning a model to reconstruct the original text.

Embeddings

MODEL : GIST Embedding v0

The model does not require any instruction for generating embeddings. This means that queries for retrieval tasks can be directly encoded without crafting instructions.
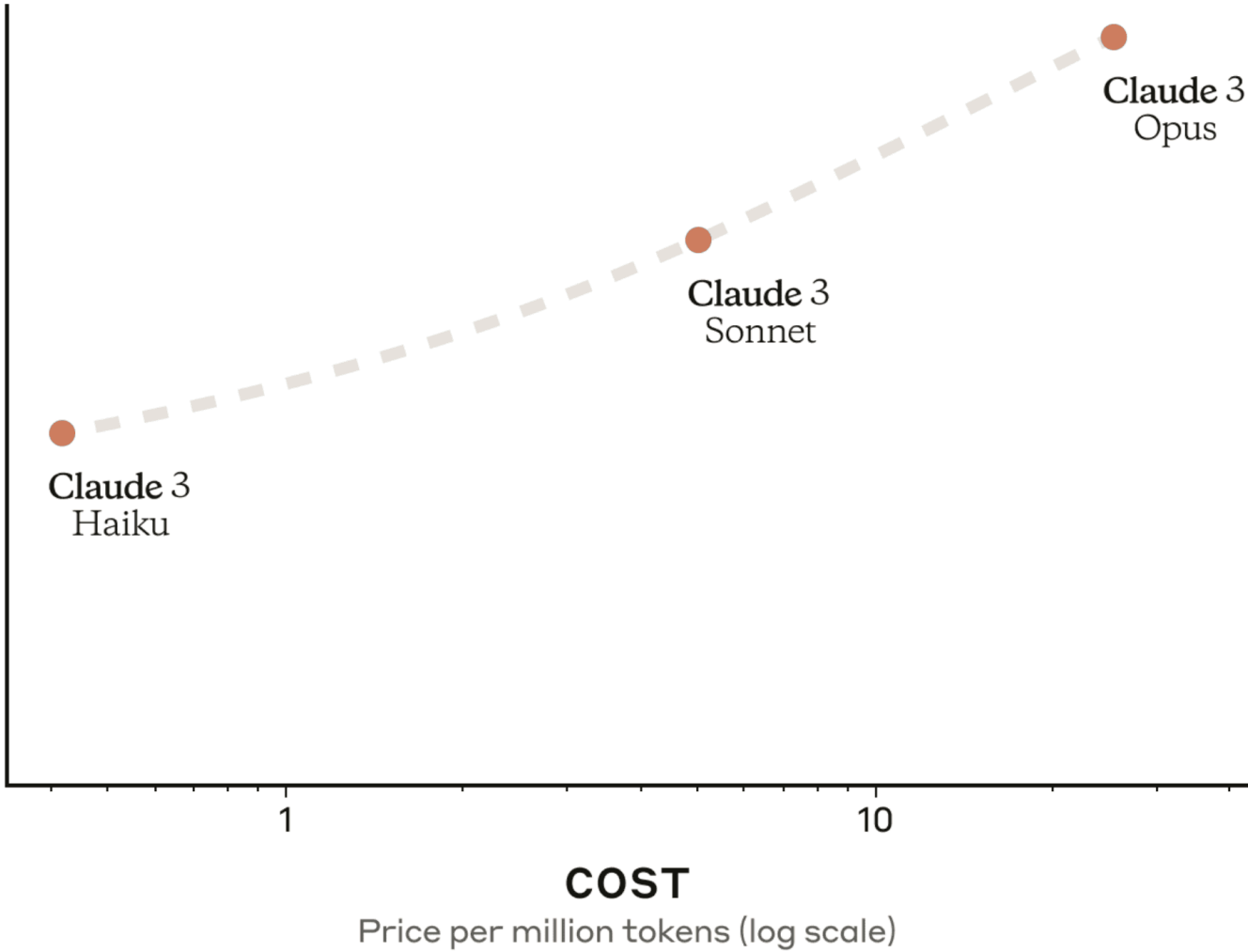
ANTHROP\C
Claude 3: A new generation of AI

(Recently released : 4 Mar, 2024) !!!

Claude is a family of state-of-the-art large language models developed by Anthropic.
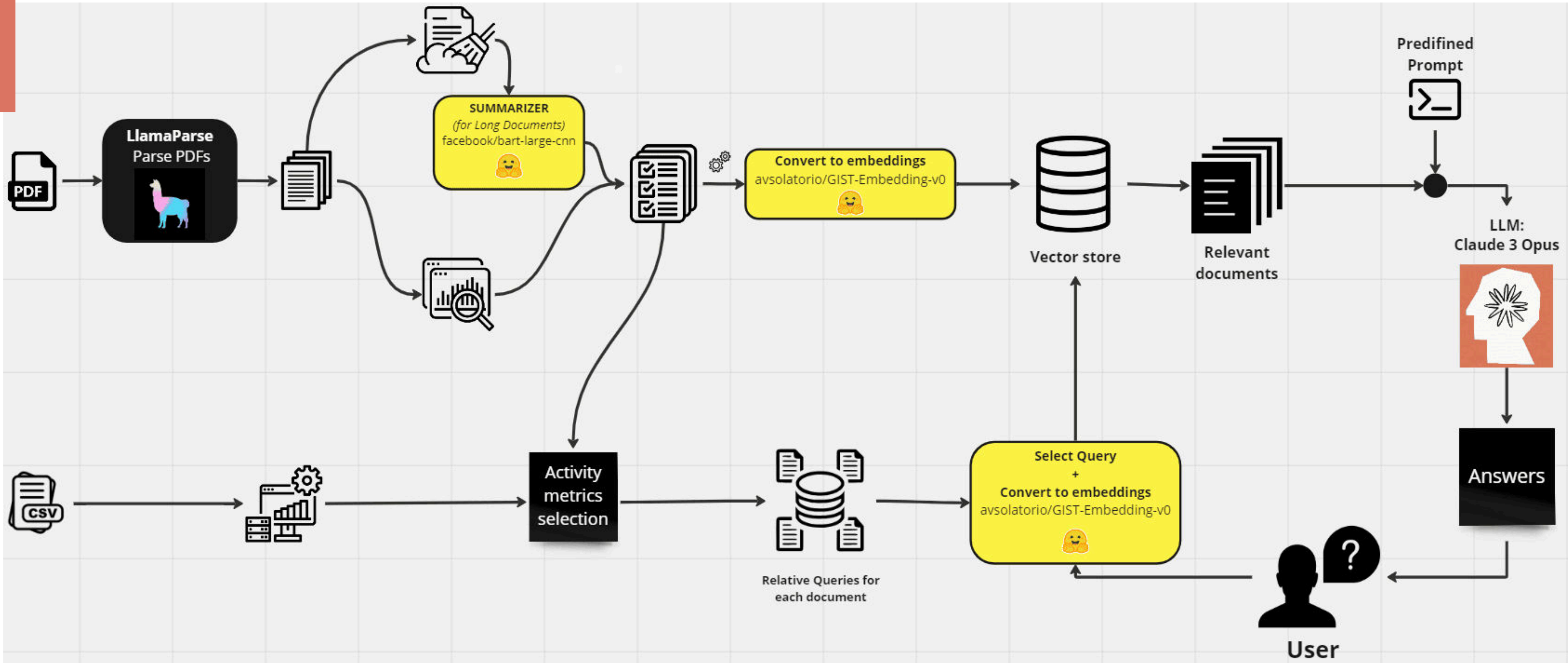
	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4	GPT-3.5	Gemini 1.0 Ultra	Gemini 1.0 Pro
Undergraduate level knowledge *MMLU*	86.8% 5 shot	79.0% 5-shot	75.2% 5-shot	86.4% 5-shot	70.0% 5-shot	83.7% 5-shot	71.8% 5-shot
Graduate level reasoning *GPQA, Diamond*	50.4% 0-shot CoT	40.4% 0-shot CoT	33.3% 0-shot CoT	35.7% 0-shot CoT	28.1% 0-shot CoT	—	—
Grade school math *GSM8K*	95.0% 0-shot CoT	92.3% 0-shot CoT	88.9% 0-shot CoT	92.0% 5-shot CoT	57.1% 5-shot	94.4% Maj1@32	86.5% Maj1@32
Math problem-solving *MATH*	60.1% 0-shot CoT	43.1% 0-shot CoT	38.9% 0-shot CoT	52.9% 4-shot	34.1% 4-shot	53.2% 4-shot	32.6% 4-shot
Multilingual math *MGSM*	90.7% 0-shot	83.5% 0-shot	75.1% 0-shot	74.5% 8-shot	—	79.0% 8-shot	63.5% 8-shot
Code *HumanEval*	84.9% 0-shot	73.0% 0-shot	75.9% 0-shot	67.0% 0-shot	48.1% 0-shot	74.4% 0-shot	67.7% 0-shot
Reasoning over text *DROP, F1 score*	83.1 3-shot	78.9 3-shot	78.4 3-shot	80.9 3-shot	64.1 3-shot	82.4 Variable shots	74.1 Variable shots
Mixed evaluations *BIG-Bench-Hard*	86.8% 3-shot CoT	82.9% 3-shot CoT	73.7% 3-shot CoT	83.1% 3-shot CoT	66.6% 3-shot CoT	83.6% 3-shot CoT	75.0% 3-shot CoT
Knowledge Q&A *ARC-Challenge*	96.4% 25-shot	93.2% 25-shot	89.2% 25-shot	96.3% 25-shot	85.2% 25-shot	—	—
Common Knowledge *HellaSwag*	95.4% 10-shot	89.0% 10-shot	85.9% 10-shot	95.3% 10-shot	85.5% 10-shot	87.8% 10-shot	84.7% 10-shot

INTELLIGENCE Benchmark scores

Claude 3 Opus

Claude 3 Sonnet

Claude 3 Haiku

COST Price per million tokens (log scale)

1   10

SOLUTION OVERVIEW

Ablation Study

TECHNIQUES	ROUND 1	ROUND 2	ROUND 3
PYPDF	✓	✓	✗
LLAMAPARSER	✗	✗	✓
WITH PREPROCESSING	✗	✓	✓
MODEL SCORE :	0.919765166	0.932159165	0.947162426

EVALUATION

Input Tokens ≈ 140,000

Output Tokens ≈ 15000

Model	Embedding	Runtime	Cost	LB	PB
Sonnet	GIST	336.10 sec	0.24$	0.944553163	0.938216382
Opus	GIST	703.91 sec	3.23$	0.947162426	0.942968968

CONCLUSION

Model deployment strategy :

- **Key Components:** Cloud-Based API Integration: Leverage the existing API access to the LLM, eliminating the need for local hosting and maintenance.
- **Web Application Interface:** Develop a user-friendly web app for uploading annual reports (PDF format).
- **Preprocessing & Data Extraction:** Integrate the preprocessing pipeline and LLM interaction code into the web app backend.
- **Data Visualization & Download:** Display extracted data in an interactive format within the web app, with download options (CSV, Excel).
- **Containerization & Scalability:** Utilize Docker for containerization, ensuring portability and facilitating deployment on cloud platforms for horizontal scaling.

For a comprehensive discussion of the deployment strategy and other aspects of this project, please refer to the accompanying report.

Thanks for your attention!

REFERENCES

https://zindi.africa/competitions/unifi-value-frameworks-pdf-lifting-competition
https://www.llamaindex.ai/blog/introducing-llamacloud-and-llamaparse-af8cedf9006b
https://huggingface.co/facebook/bart-large-cnn
https://docs.anthropic.com/claude/docs/models-overview
https://www.anthropic.com/news/claude-3-family