TEAM MEMBERS:
AARSHIT JOLAPARA
TAHA BHATIA
MAHIN GANDHARV
KATHAN KATHROTIYA
AAMIR KHARODAWALA

TEAM : FINETUNERS

MINED HACKATHON 2024

COMPANY NAME : CACTUS COMMUNICATIONS

TRACK : CHATWITHANYSCIENTIFICDOCUMENT

IDEATION PROCESS

01

Design and implement an intelligent document parser capable of processing multiple file formats, ensuring compatibility with PDFs, .docx, LaTeX (.tex), and .ppt files.

02

Implement robust preprocessing techniques to standardize document inputs, facilitating smooth integration with LLMs while minimizing noise and maximizing signal extraction.

03

Develop connectors for popular LLMs such as GPT, Llama, Mistral, and Claude, enabling seamless communication between the parsed documents and the language models.

IDEATION PROCESS

04

Fine-tune the document processing pipeline to remove hallucinations by optimizing parameters, model architectures, and training methodologies, ensuring reliable and accurate document analysis in LLMs.

05

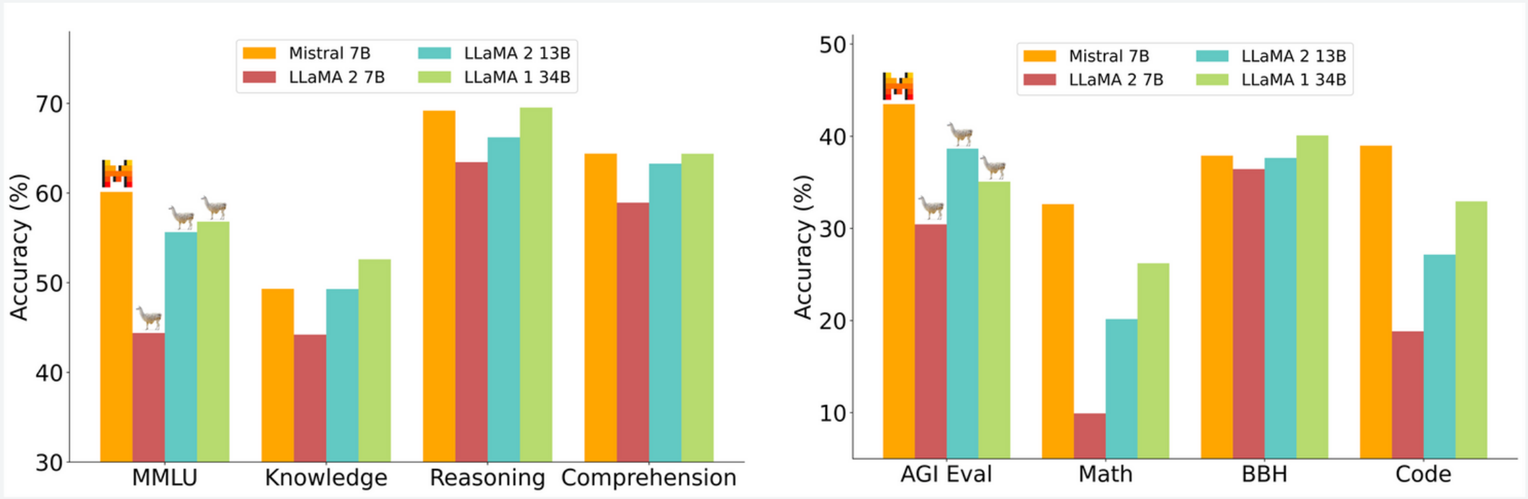
Create an intuitive chat interface that allows users to upload documents, pose questions, and receive accurate answers generated by connected LLMs, thereby enhancing user interaction and accessibility.

MODEL IDENTIFICATION

Model Architecture

Mistral-7B-v0.1 is a transformer model, with the following architecture choices:

- Grouped-Query Attention for faster inference
- Sliding-Window Attention to handle longer sequences at smaller cost
- Byte-fallback BPE tokenizer
- Mistral 7B is easy to fine-tune on any task.



Model	Modality	MMLU	HellaSwag	WinoGrande	PIQA	Arc-e	Arc-c	NQ	TriviaQA	HumanEval	MBPP	MATH	GSM8K
LLaMA 2 7B	Pretrained	44.4%	77.1%	69.5%	77.9%	68.7%	43.2%	24.7%	63.8%	11.6%	26.1%	3.9%	16.0%
LLaMA 2 13B	Pretrained	55.6%	**80.7%**	72.9%	80.8%	75.2%	48.8%	**29.0%**	**69.6%**	18.9%	35.4%	6.0%	34.3%
Code LLaMA 7B	Finetuned	36.9%	62.9%	62.3%	72.8%	59.4%	34.5%	11.0%	34.9%	**31.1%**	**52.5%**	5.2%	20.8%
Mistral 7B	Pretrained	**60.1%**	**81.3%**	**75.3%**	**83.0%**	**80.0%**	**55.5%**	**28.8%**	**69.9%**	**30.5%**	47.5%	**13.1%**	**52.1%**

ADVANTAGES

01 MODULAR STRUCTURE FOR SCALABILITY AND COLLABORATION.

02 EFFICIENT TEXT CHUNKING WITH OVERLAPPING FOR LARGE DATASET PROCESSING.

03 OPTIMIZATION FEATURES LIKE **QUANTIZATION** AND GPU ACCELERATION FOR PERFORMANCE BOOST.

04 STRUCTURED PIPELINE ENSURES STREAMLINED DATA FLOW AND ERROR HANDLING.

05 FLEXIBILITY WITH **CUSTOM PROMPTS** AND RETRIEVAL STRATEGY ADJUSTMENTS.

06 **INSTRUCTION FINE TUNED** MODEL FOR BETTER QUERY RESPONSES

07 SUPPORTS **MULTIPLE FILE FORMATS** (.PDF, .DOCX, .PPTX)

FUTURE WORK

01 EXPERIMENT WITH FINE-TUNING STRATEGIES TO REDUCE HALLUCINATIONS.

02 DEVELOP AN INTUITIVE USER INTERFACE.

03 INCORPORATE PRIVACY-PRESERVING MEASURES.

04 EXTEND SUPPORT FOR MULTIPLE DOCUMENTS.

THANK YOU