Agentic AI for Medical Imaging and Diagnosis Support

Pratik Bhangale

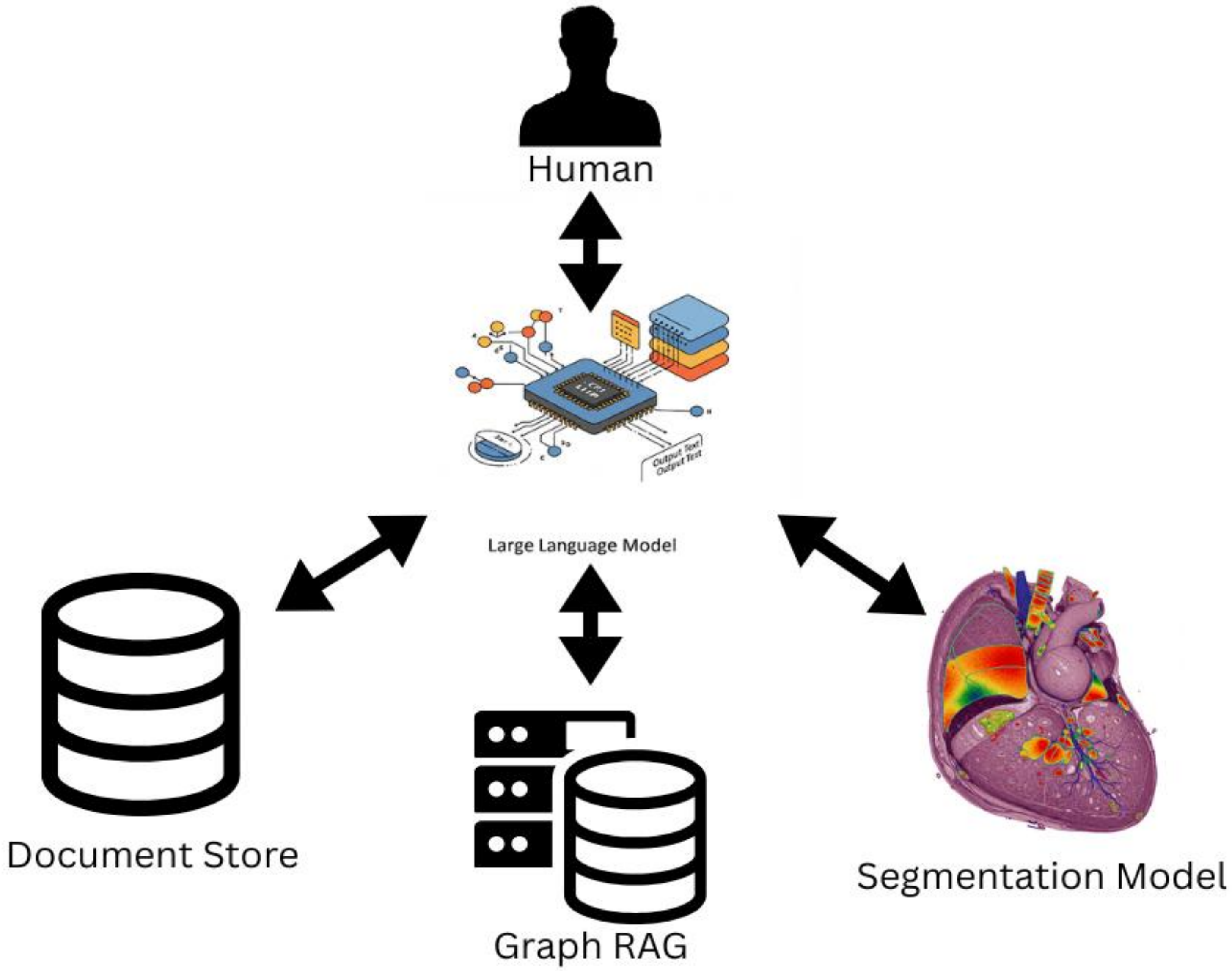
Introduction

Research objectives:

- Develop an AI system for medical image segmentation
- Integrate LLMs for knowledge retrieval and decision support

Research questions:

- Can Unet and LLMs improve medical image segmentation accuracy and ease of use?
- Can the proposed system enhance decision-making for healthcare professionals?



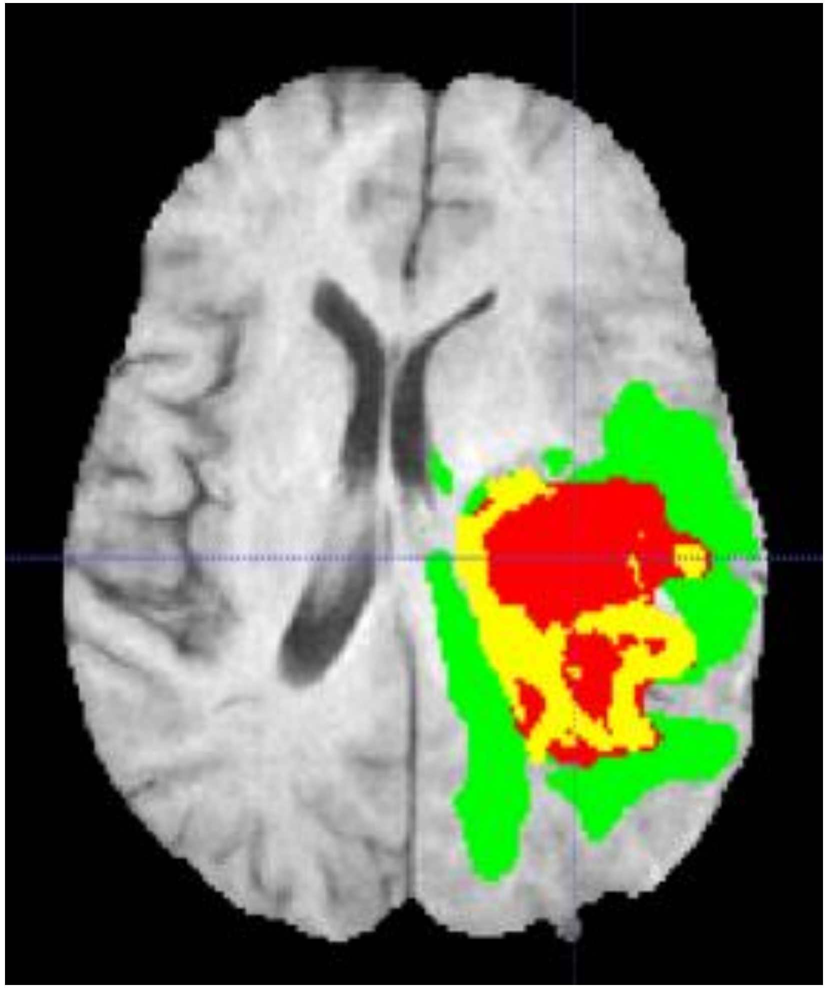
Research Hypothesis

Research Hypotheses:

- Null Hypothesis (H0): No significant difference in segmentation accuracy between proposed AI system and traditional methods.
- Alternative Hypothesis 1 (H1): Proposed AI system significantly improves segmentation accuracy.
- Alternative Hypothesis 2 (H2): Proposed AI system improves segmentation accuracy and provides better knowledge retrieval and decision support.

Metrics for Hypothesis Testing:

- Segmentation Accuracy: Dice coefficient, IoU, Hausdorff distance
- Knowledge Retrieval: F1-score, BLEU score, Meteor score
- Decision Support: User feedback, Clinical impact



Brain Tumor Segmentation

Proposed Core Algorithm (Segmentation)

Architecture

- <u>Backbone</u>: UNet (U-Net skip connections)
- <u>Attention</u>: Spatial/channel attention gates (focus on ROI)

Dataset: Kaggle_3m (CT/MRI scans)

- <u>Preprocess</u>: Normalize intensity; augment with flips, rotations; patch extraction, and random blur.
- <u>Mask Format</u>: Pixel-wise annotations (e.g., tumors, organs)

Training Phase

- <u>Loss</u>: Hybrid (Dice + BCE). Focal Loss if necessary.
- <u>Hyperparameters</u>: LR = 3e-4 (AdamW), Batch size = 16

Attention Mechanism

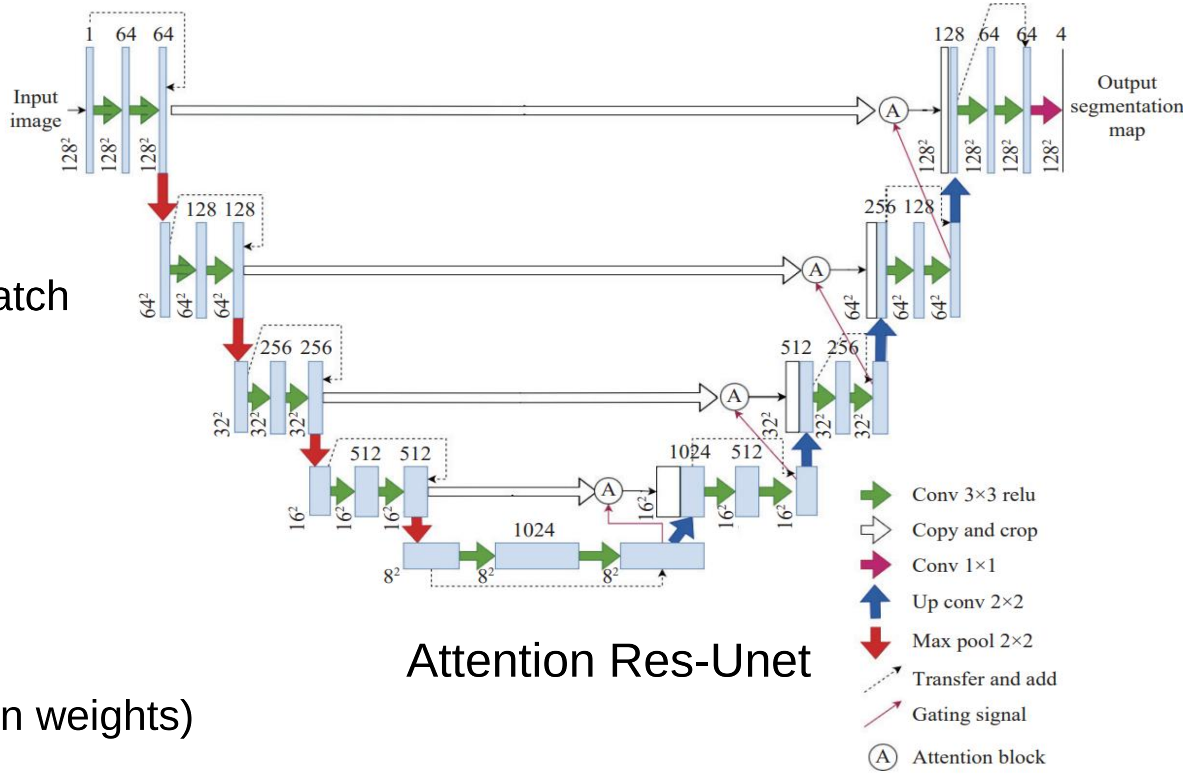
- <u>Guidance</u>: Multi-scale feature fusion (skip connections + attention weights)

Training Strategy

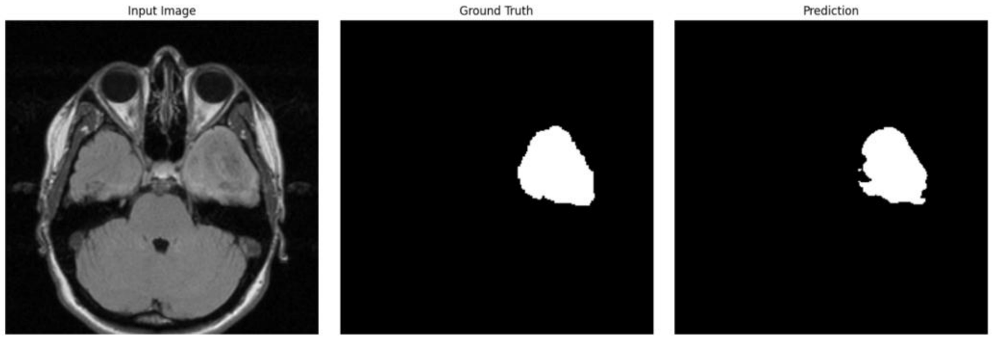
- <u>Techniques</u>: Mixed precision (FP16) + gradient clipping
- <u>Regularization</u>: Early stopping (Dice plateau), Weight decay (L2)

Evaluation

- <u>Metrics</u>: Dice score, Binary Cross Entropy, Focal Loss.
- <u>Goal</u>: Robust segmentation with minimal false positives



Attention Res-Unet



Kaggle_3m Dataset

Proposed Core Algorithm (LLM Distillation)

Architecture
- Student: LLaMA 3.2 3B (fine-tuned → distilled).
- Teacher: Frozen LLaMA 3.1 8B (logits/CoT guidance).

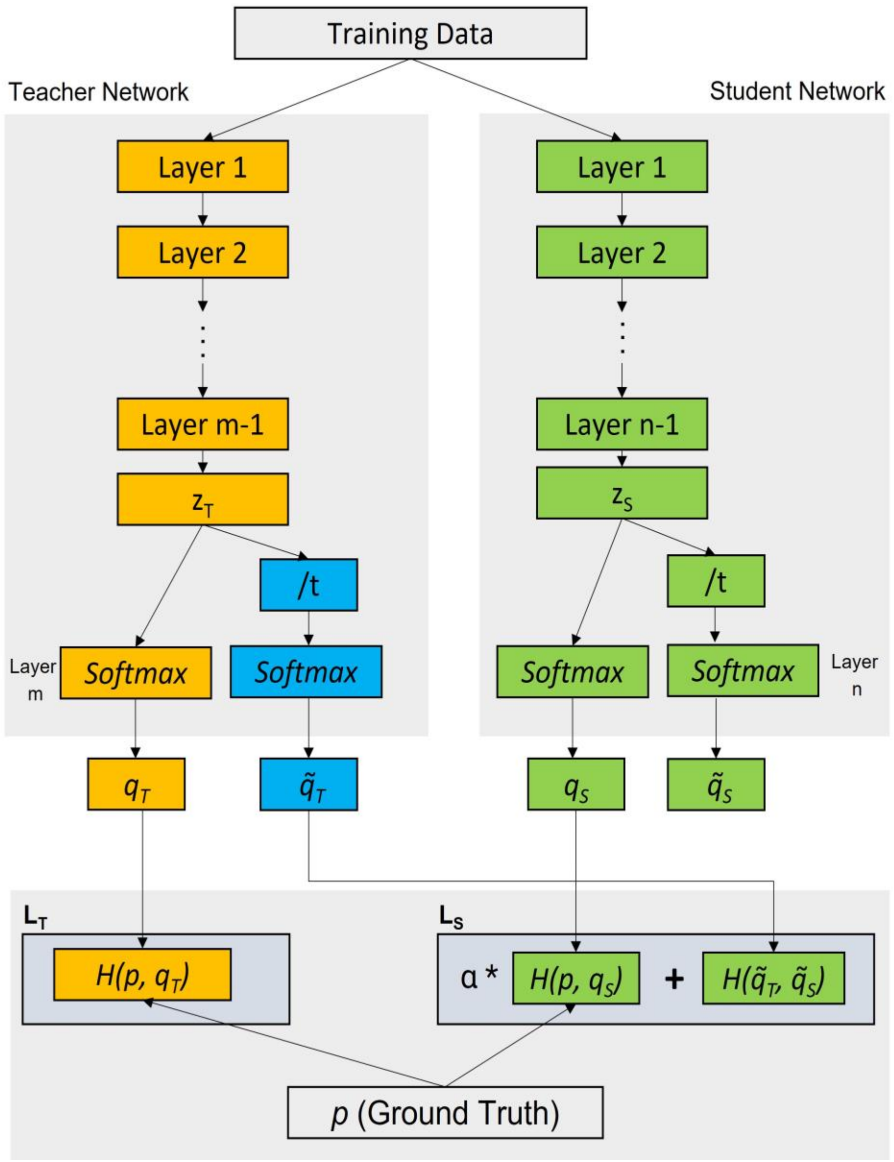
Dataset: HPAI-BSC/MMLU-medical-cot-llama31
- Preprocess: QA → instruction prompts
- Tokenize: LLaMA-3 tokenizer (pad/truncate).

Fine-Tuning Phase
- Loss: Causal LM (supervised).
- Hyperparams: LR=1e-5, memory-optimized batch size.

Distillation Phase
- Hybrid Loss: Cross-entropy (ground truth) + KL-divergence.
- Dynamic Weighting: SFT (70%) + Teacher (30%).

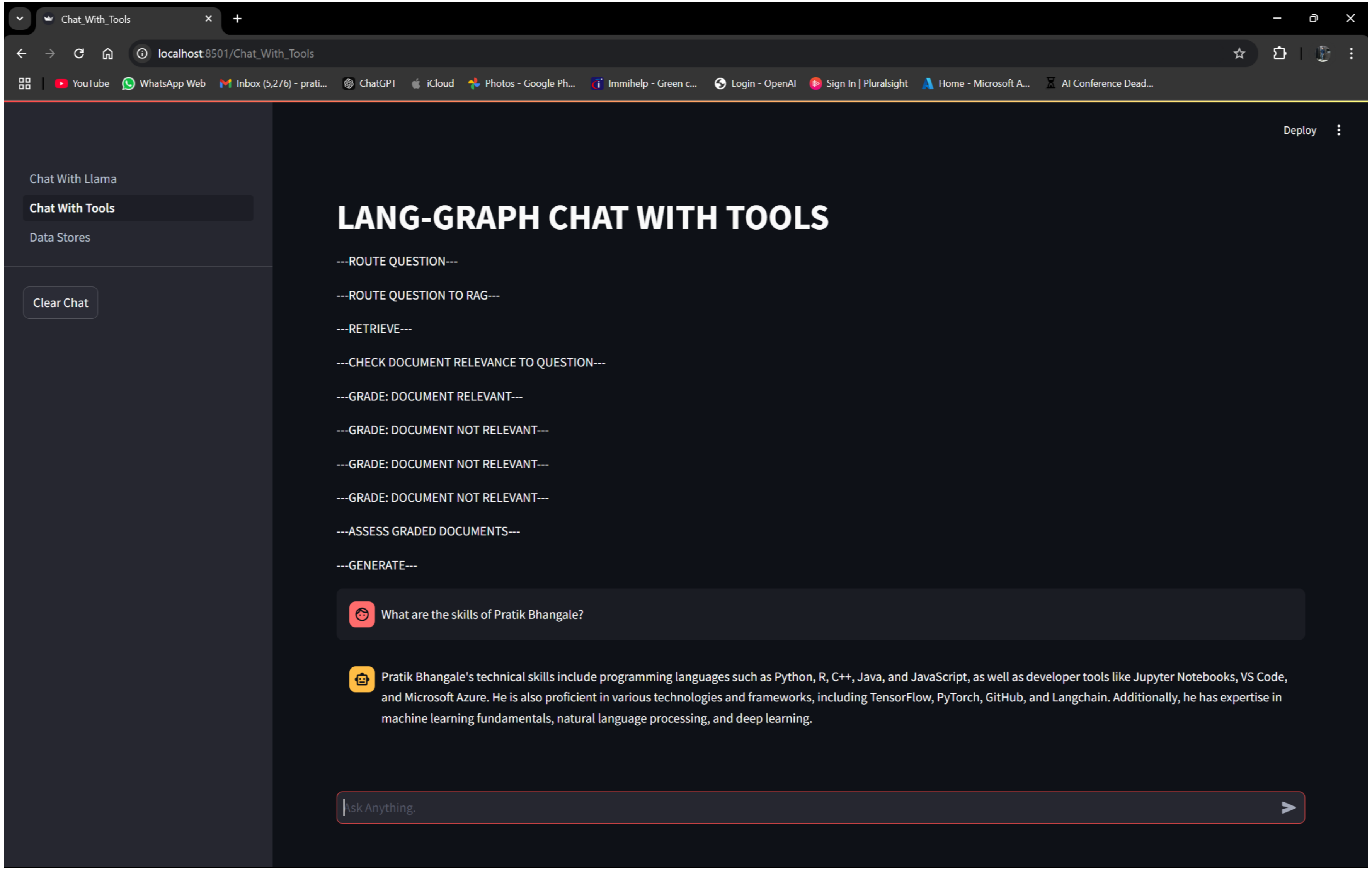
Training Strategy
- Progressive distillation (SFT → distillation focus).
- FP16/Flash Attention + dropout/weight decay.

Evaluation
- Metrics: Perplexity and Meteor.
- Goal: 3B model with 8B-level medical reasoning, lower inference cost.



Combined System

- LangGraph + Streamlit: Interactive website integrating distilled LLaMA model with Attention ResUNet model and RAG.
- RAG databases: Case studies + Graph RAG for contextual retrieval.
- Attention ResUNet model: Tool for LLM agent to Analyse images for Brain Tumors.



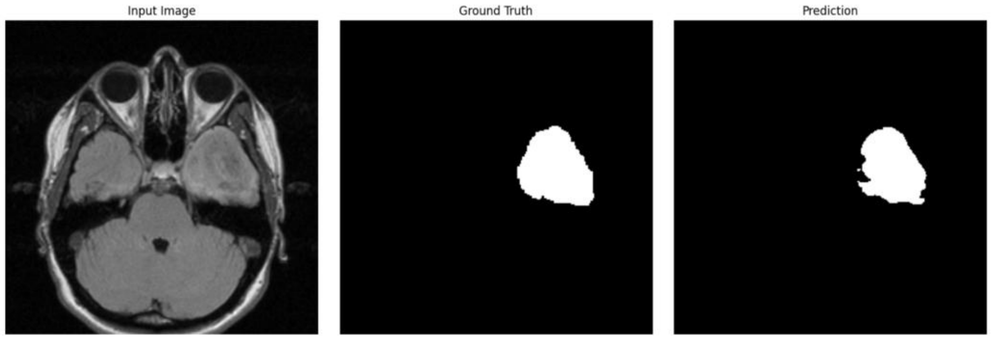
Baseline Results

Unet Baseline

Train Loss: 0.0383

Val Loss: 0.1056

Loss = (0.5*BCE)+(0.5*Dice)

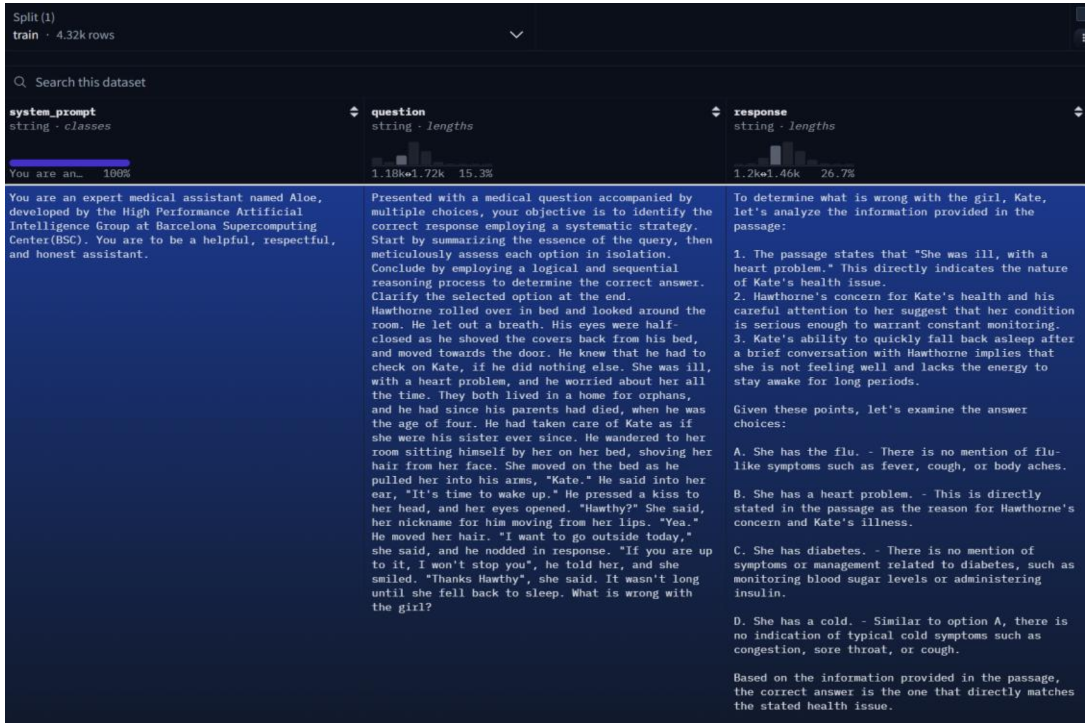


Kaggle_3m

Distillation Baseline

METEOR Score: 32.74%

Perplexity: 12.89



HPAI-BSC/MMLU-medical-cot-llama31

Conclusion

- Successful AI Integration: The proposed system combines Attention Res-UNet for precise medical image segmentation and a distilled LLaMA model for enhanced knowledge retrieval and decision support.
- Improved Accuracy & Efficiency: Segmentation: Achieved strong Dice and IoU scores with minimal false positives.
- Knowledge Retrieval: Competitive METEOR and Perplexity scores, enabling better clinical decision-making.
- Potential Clinical Impact: Supports healthcare professionals in diagnosis. Reduces manual workload while improving diagnostic reliability.
- Future Work: Expand dataset diversity for better generalization.Optimize model efficiency for real-time applications.Validate in real-world clinical settings.

Thank You.