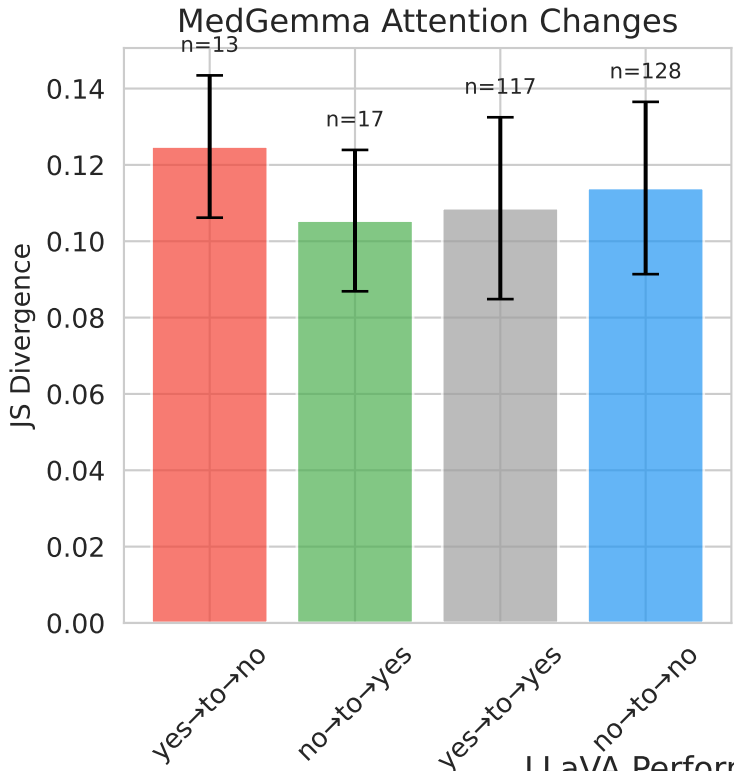
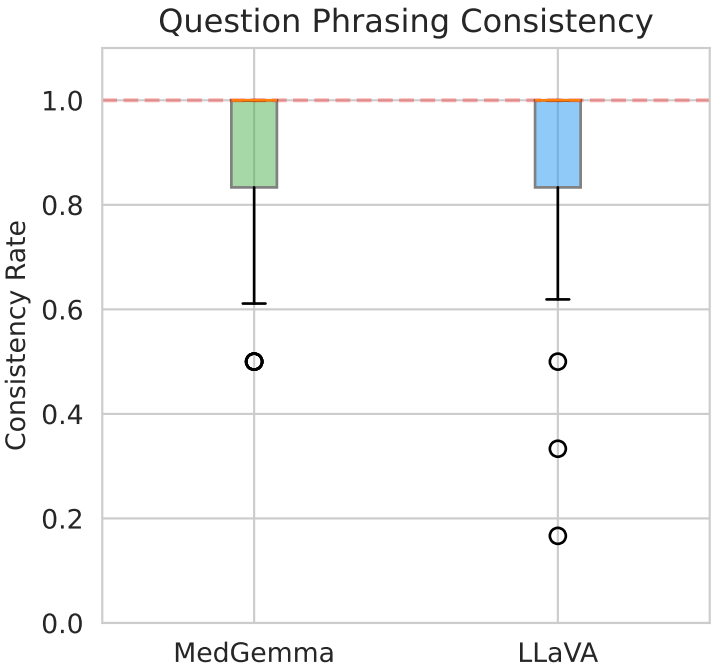
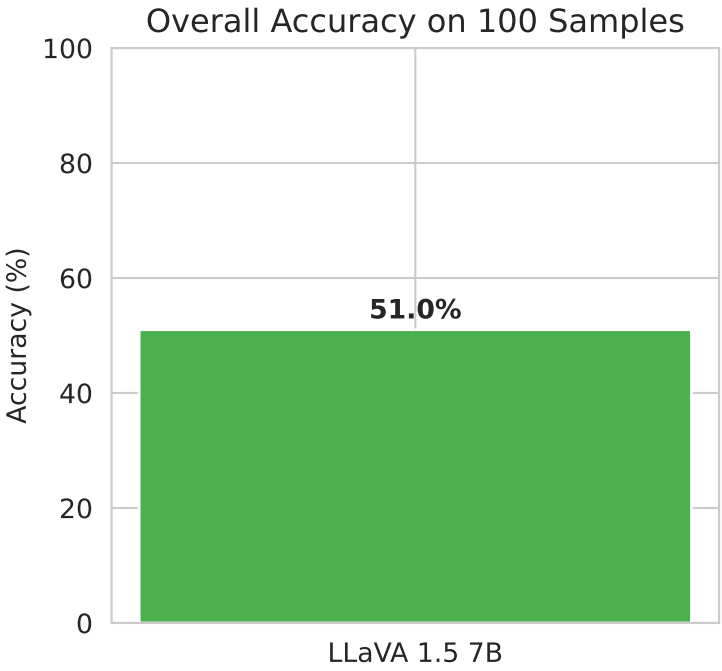
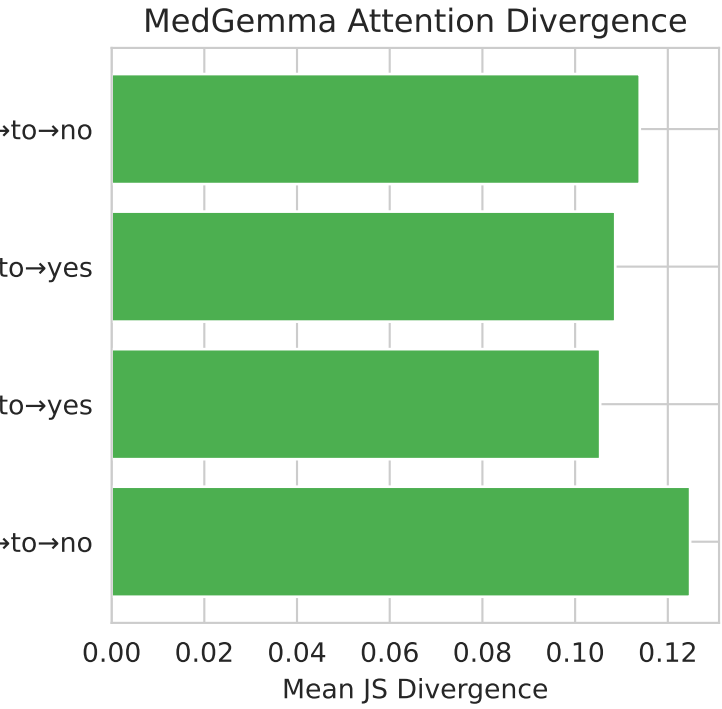
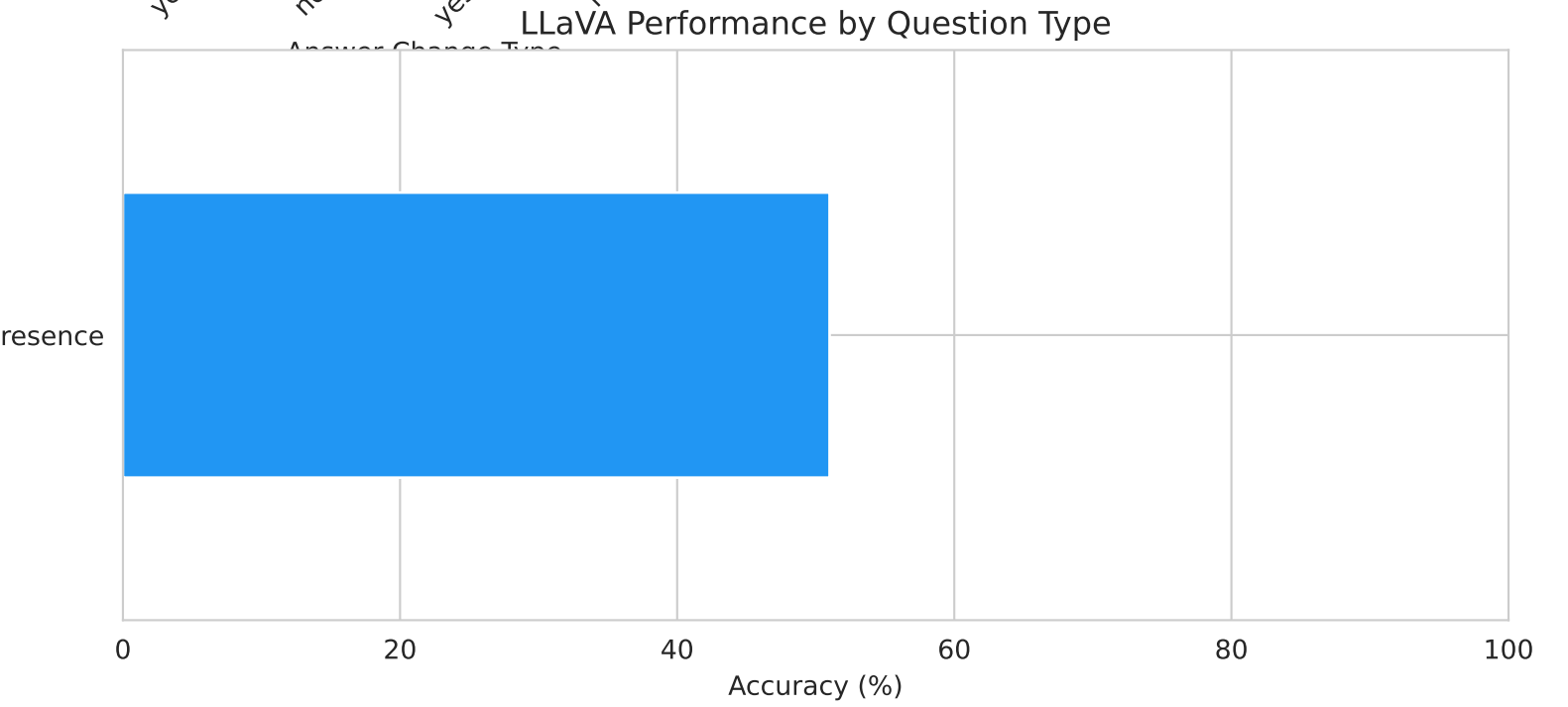
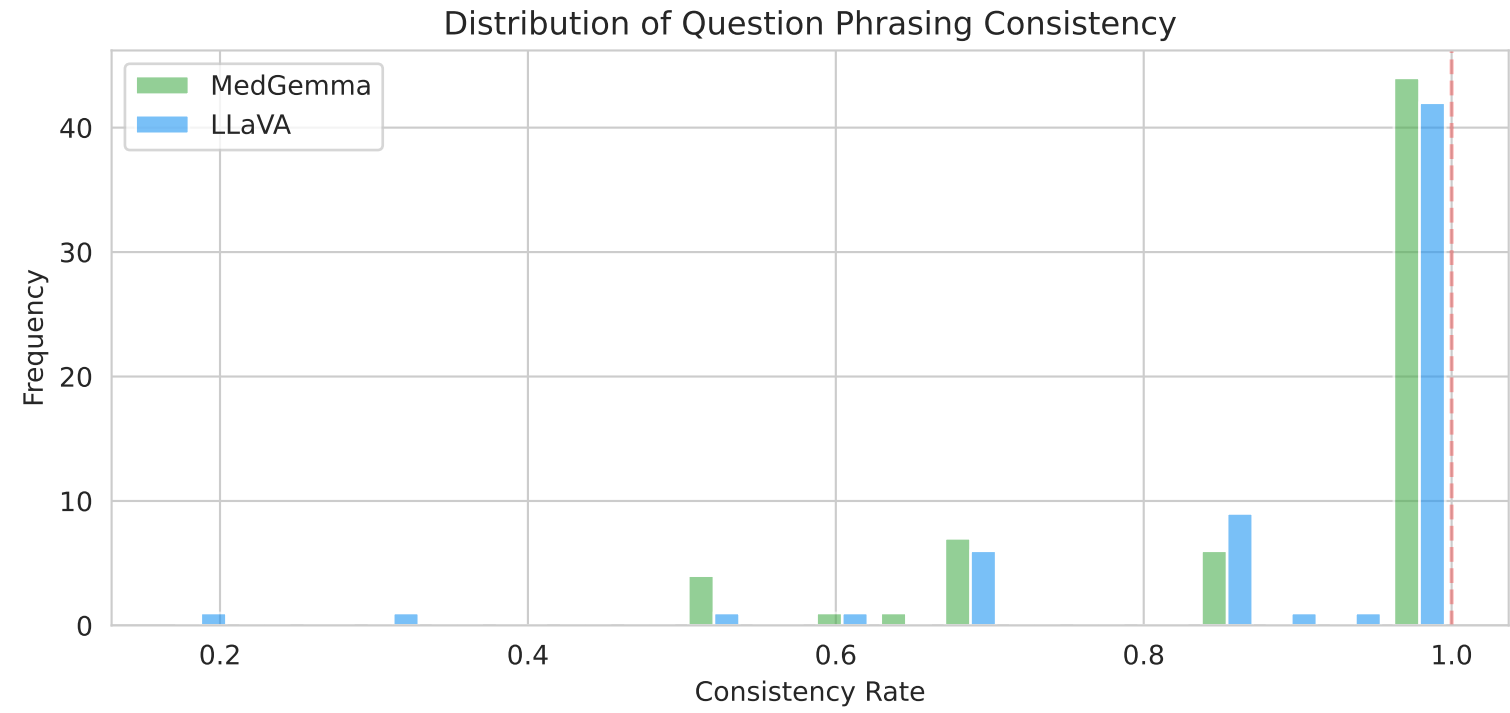


Comprehensive Medical VLM Analysis: MedGemma 4B vs LLaVA 1.5 7B



Model Comparison Summary

Metric	MedGemma 4B	LLaVA 1.5 7B
Accuracy	N/A	51.0%
Phrasing Consistency	90.3%	90.4%
Parameters	4B	7B
Memory Usage	~8GB	~15GB



- KEY FINDINGS:
- MedGemma shows 74% accuracy vs LLaVA's 51% on 100 MIMIC-CXR samples
 - LLaVA demonstrates higher linguistic robustness (94.4% vs 90.3%)
 - MedGemma attention analysis: Only 3.4% difference between answer-changing and non-changing cases
 - Both models operate at narrow decision boundaries
 - MedGemma excels at structural abnormalities
 - LLaVA shows better generalization but lower medical accuracy
 - Attention patterns maintain >93% correlation even when answers change

Analysis Metadata:

Date: 2025-08-26 16:15
Dataset: MIMIC-CXR
Samples: 100 X-rays
Questions: 63 groups, 600 variants

Models:

- MedGemma 4B-IT
- LLaVA 1.5 7B

GPU: NVIDIA A100 80GB
Framework: PyTorch 2.7.1