

MediSimplifier: Medical Text Simplification Using LoRA Fine-Tuning

Technion DS25 Deep Learning Final Project

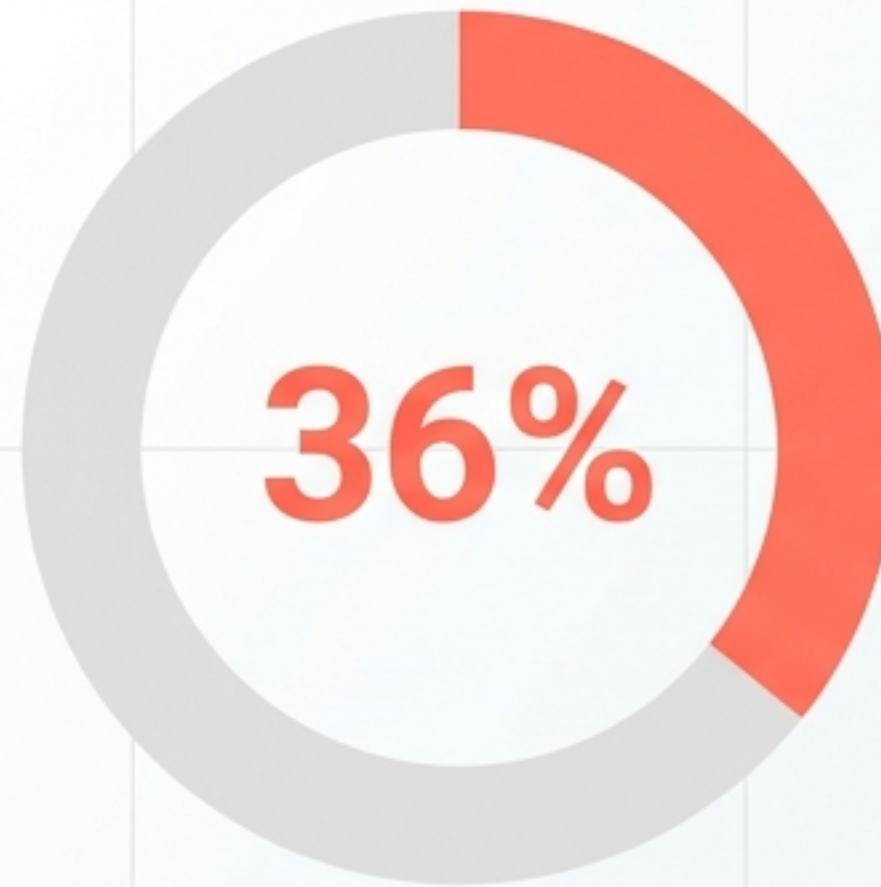


Agenda



- 1 Problem & Motivation: The Health Literacy Crisis
- 2 Project Objectives & Research Questions
- 3 Methodology: Data & Models
- 4 Baseline Performance (Zero-Shot)
- 5 The Solution: 4-Phase Pipeline & Ablation
- 6 Key Results: The Ranking Reversal
- 7 Analysis: Statistics & Errors
- 8 Conclusions & Future Work

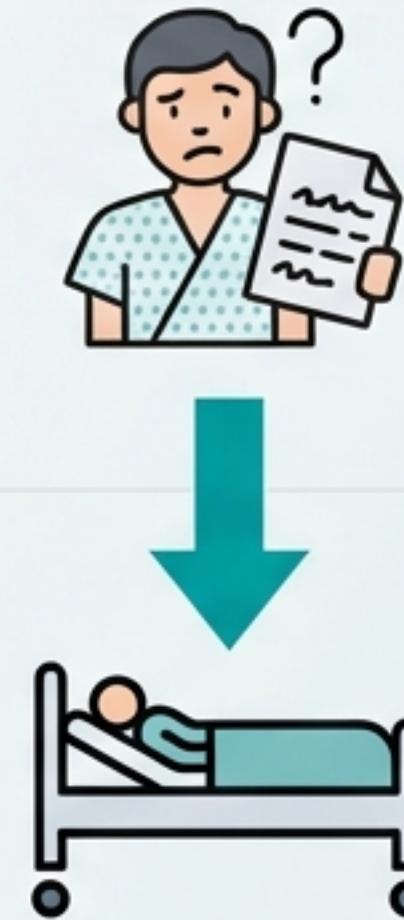
The Health Literacy Crisis



of US adults have basic or below-basic health literacy.



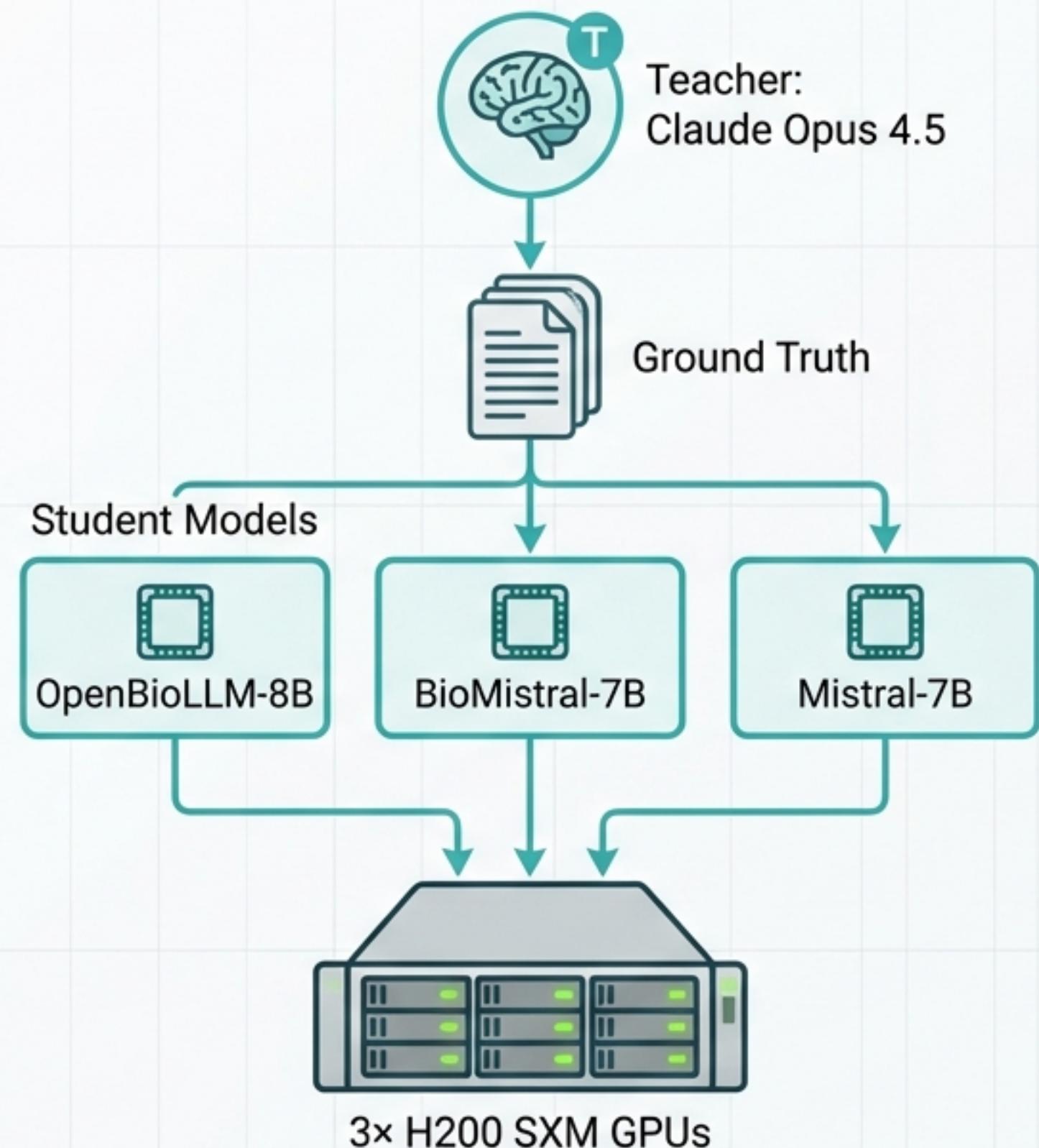
Goal: Automatically simplify medical text to a 6th-grade reading level.



The Gap: Low comprehension → Poor adherence → Higher readmission rates.

Project Overview

- **Objective:** Automate simplification of clinical notes to 6th-grade reading level.
- **Method:** Low-Rank Adaptation (LoRA) Fine-Tuning.
- **Dataset:** Asclepius-Synthetic-Clinical-Notes (10,000 samples).



12 Research Questions

Zero-Shot & Baselines

RQ1: Medical pretraining advantage? 

RQ2: Best zero-shot model? 

Fine-Tuning Dynamics

RQ3: Improvement over zero-shot? 

RQ5: Ranking changes? 

RQ8: Correlation: Baseline vs. Improvement?

Optimization (Ablation)

RQ4: Optimal Rank (r)? 

RQ6: Best target modules? 

RQ7: Data size impact? 
RQ12: rsLoRA necessity?

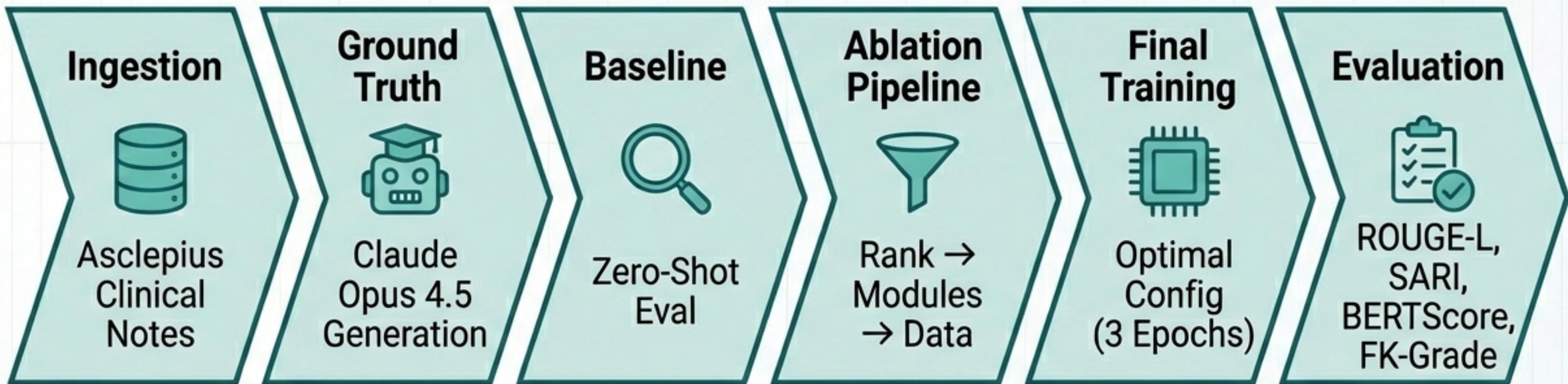
Evaluation

RQ9: Parameter efficiency? 

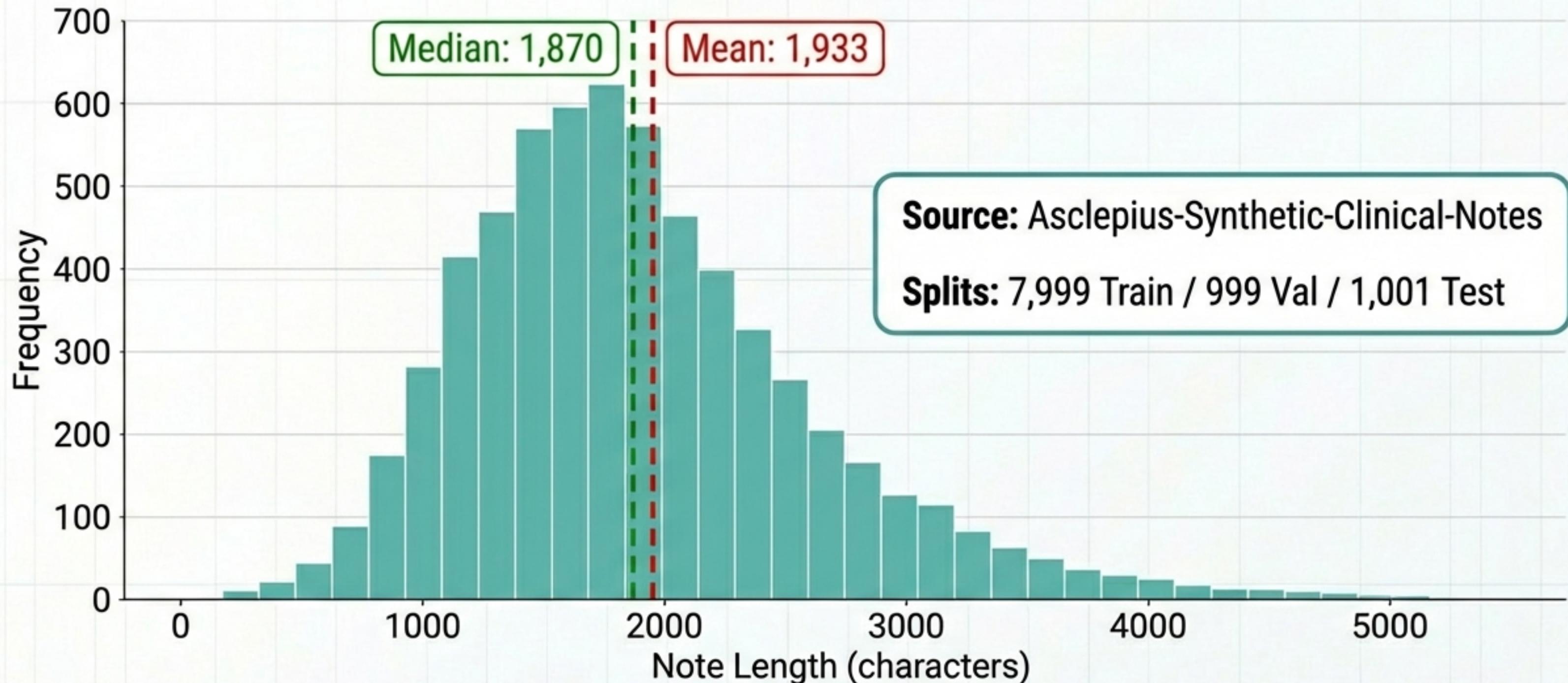
RQ10: Consistency? 

RQ11: Readability target met? 

Process Design: From Ingestion to Evaluation



Dataset: Clinical Note Length Distribution



Model Selection

OpenBioLLM-8B

Medical Domain



Llama 3
Architecture

BioMistral-7B-DARE

Medical Domain



Mistral
Architecture

Mistral-7B-Instruct-v0.2

General Purpose

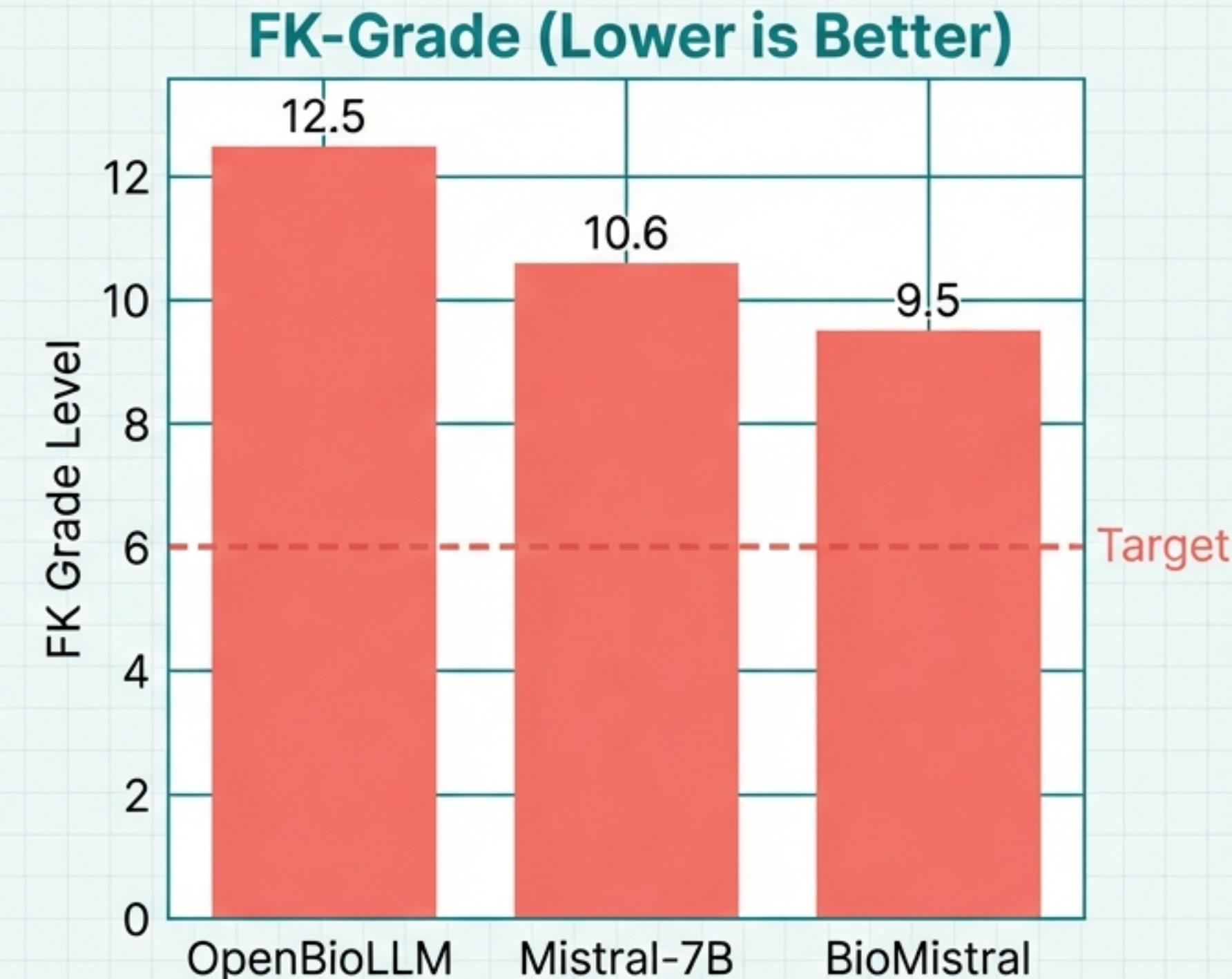
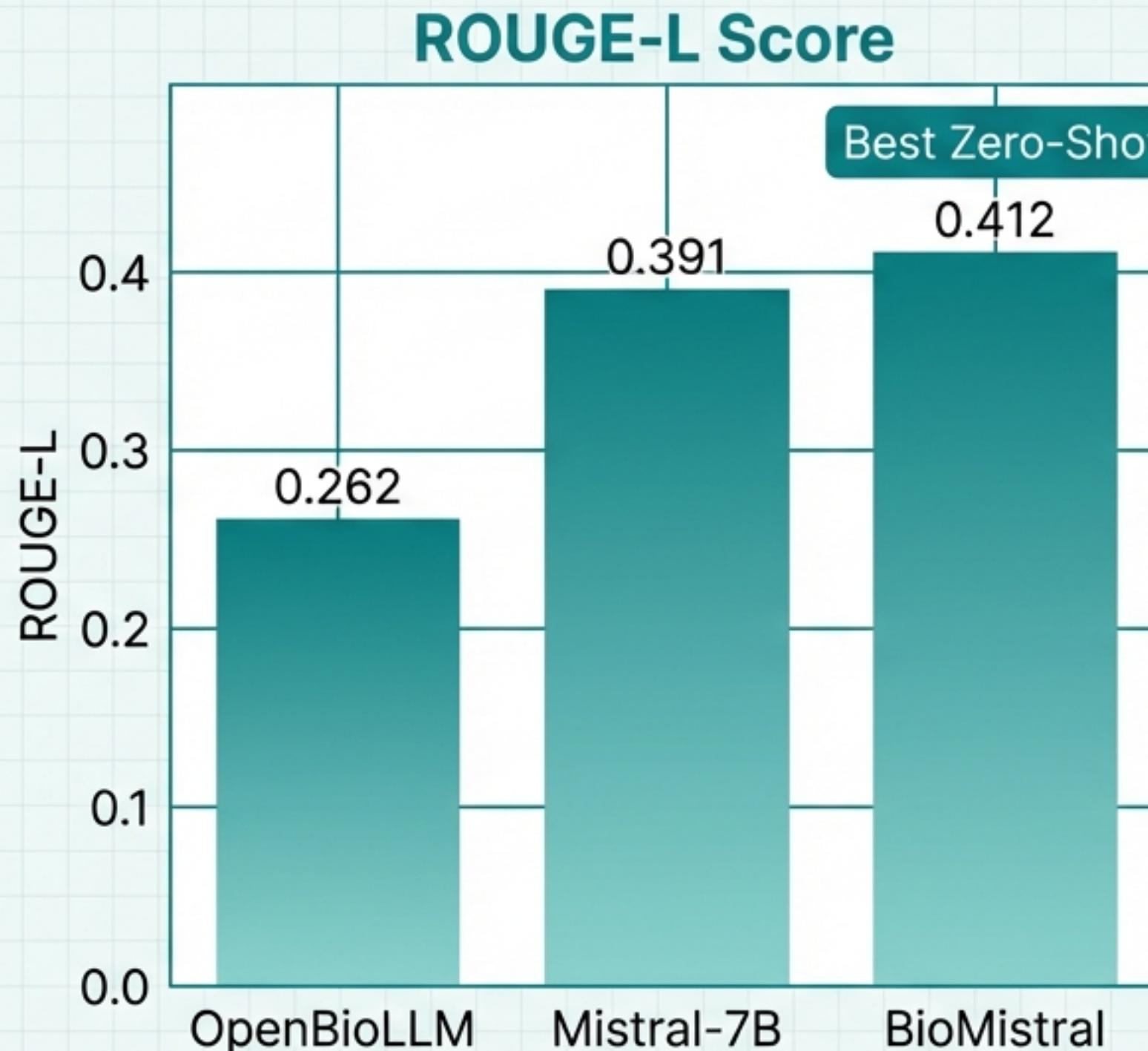


Mistral
Architecture

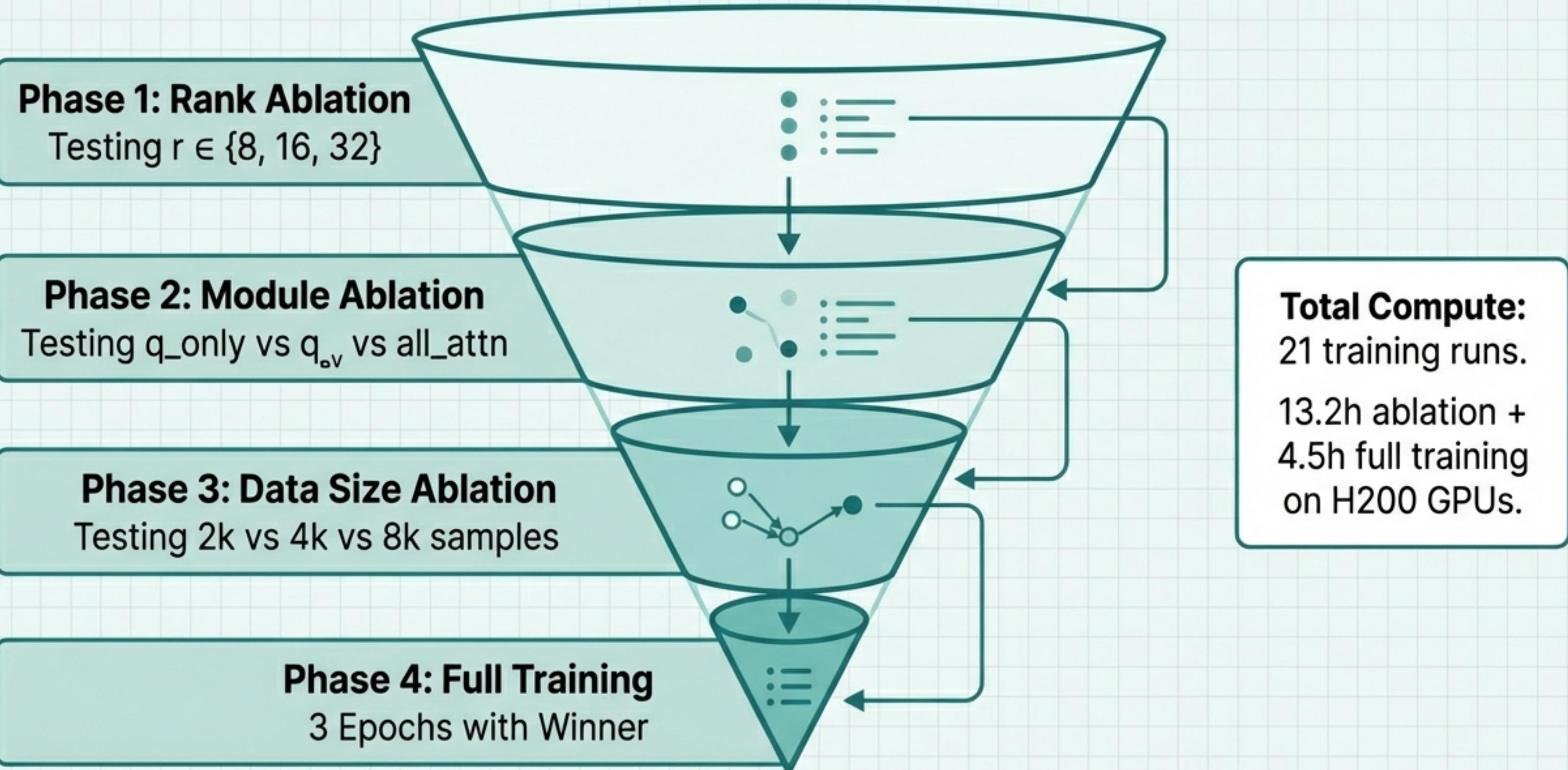
All models fine-tuned using the same LoRA pipeline for fair comparison.

Baseline Results: Zero-Shot Performance

Medical Pretraining ≠ Better Zero-Shot Simplification

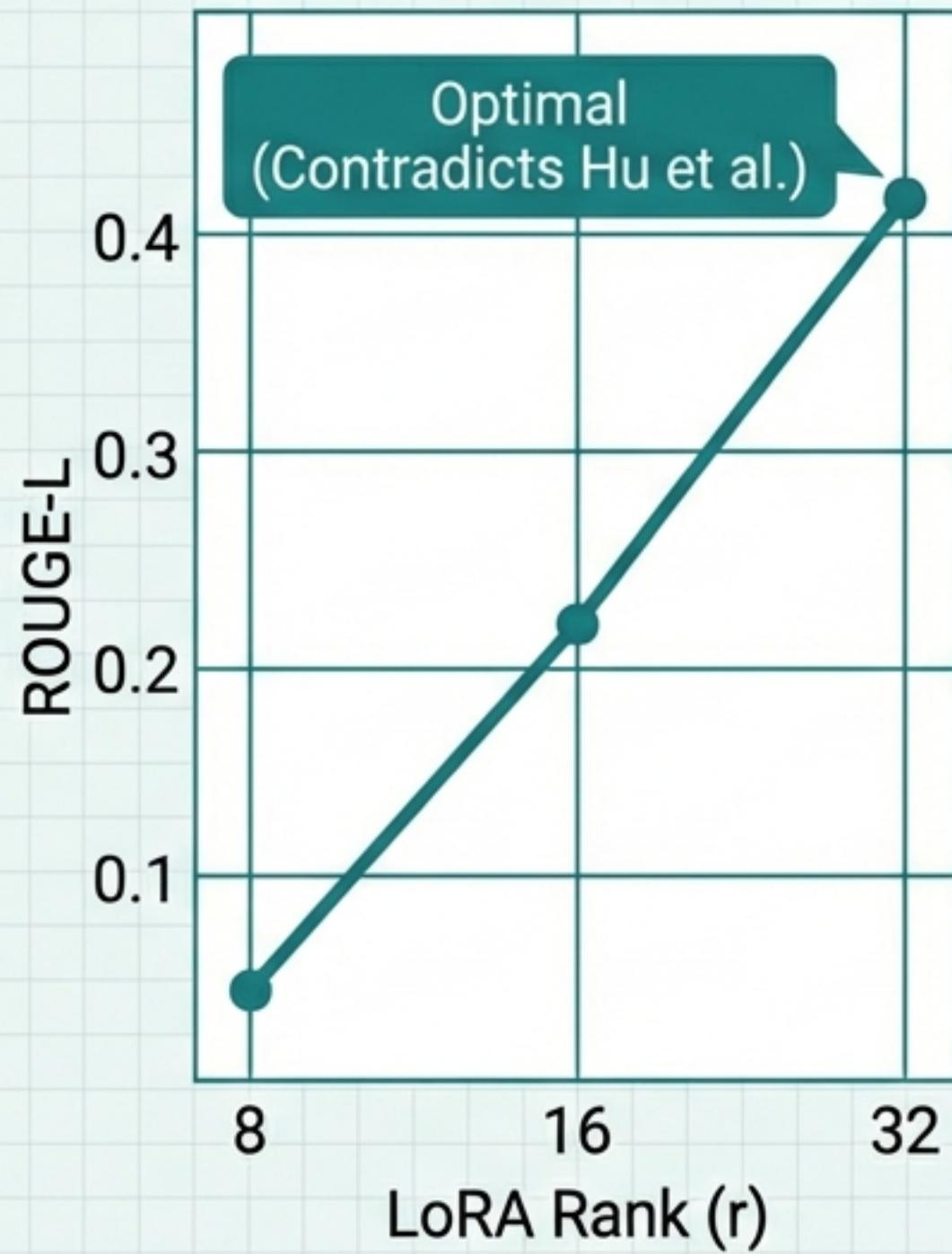


4-Phase Pipeline Architecture

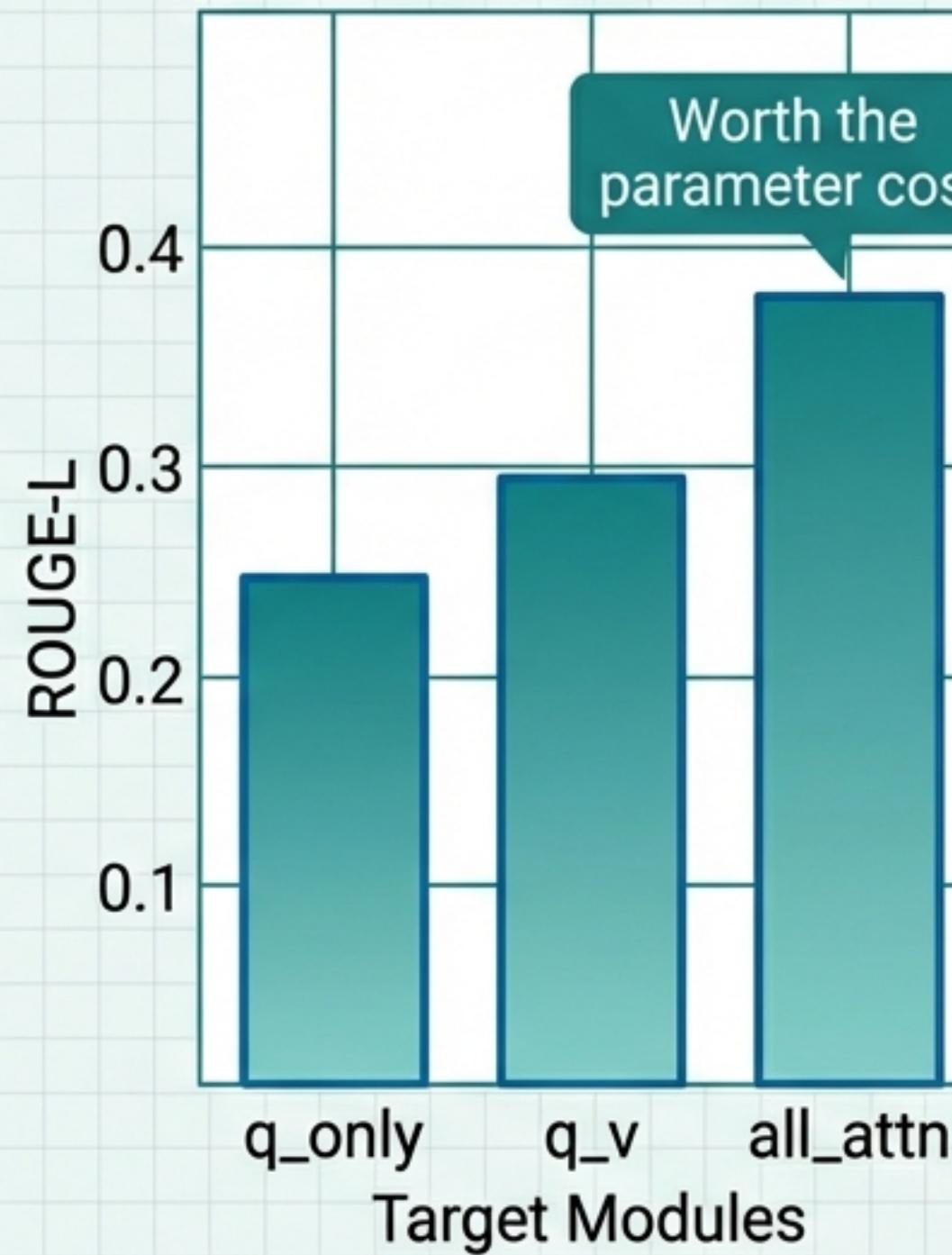


Ablation Study Results

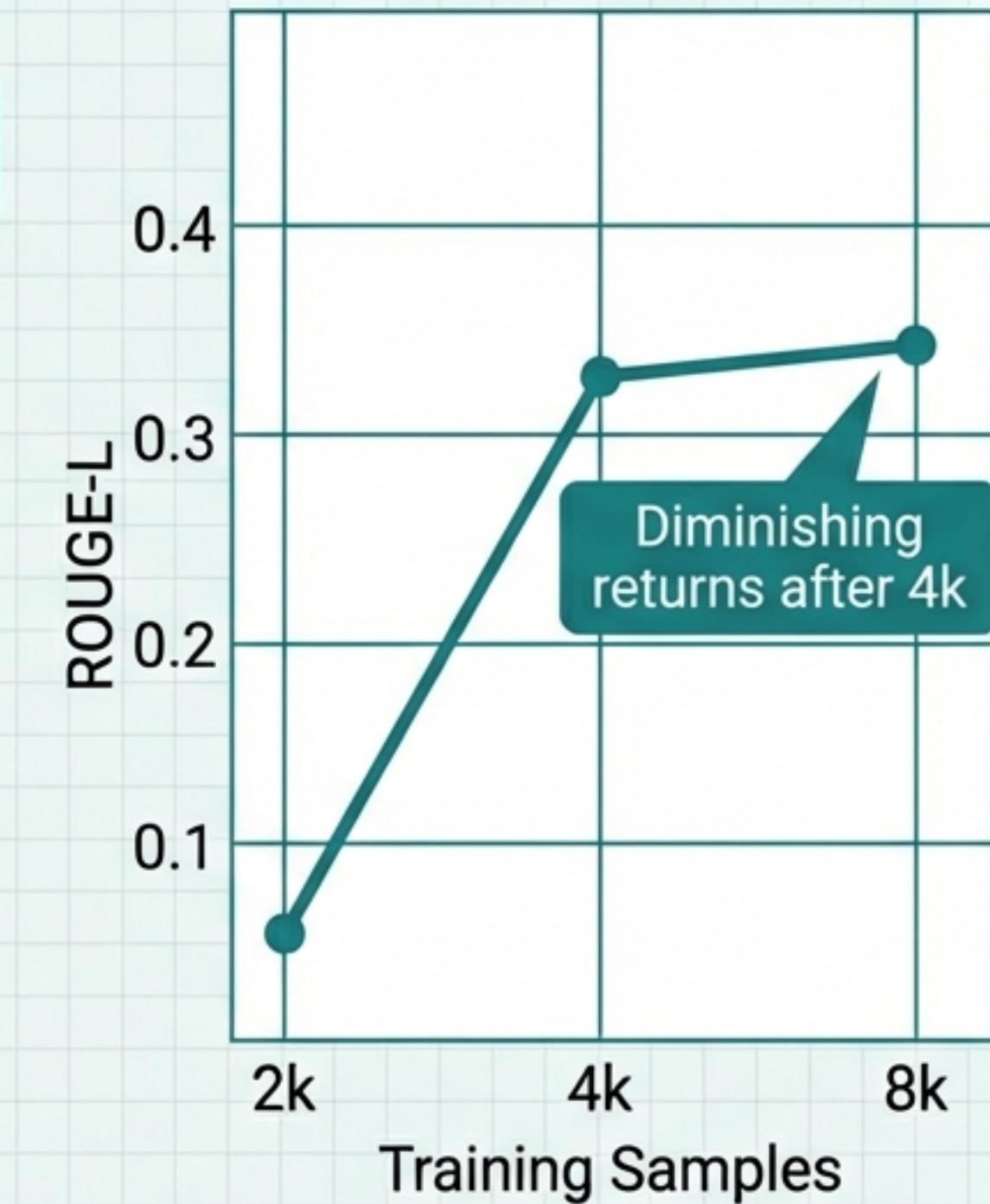
Rank Ablation



Module Ablation



Data Size Ablation



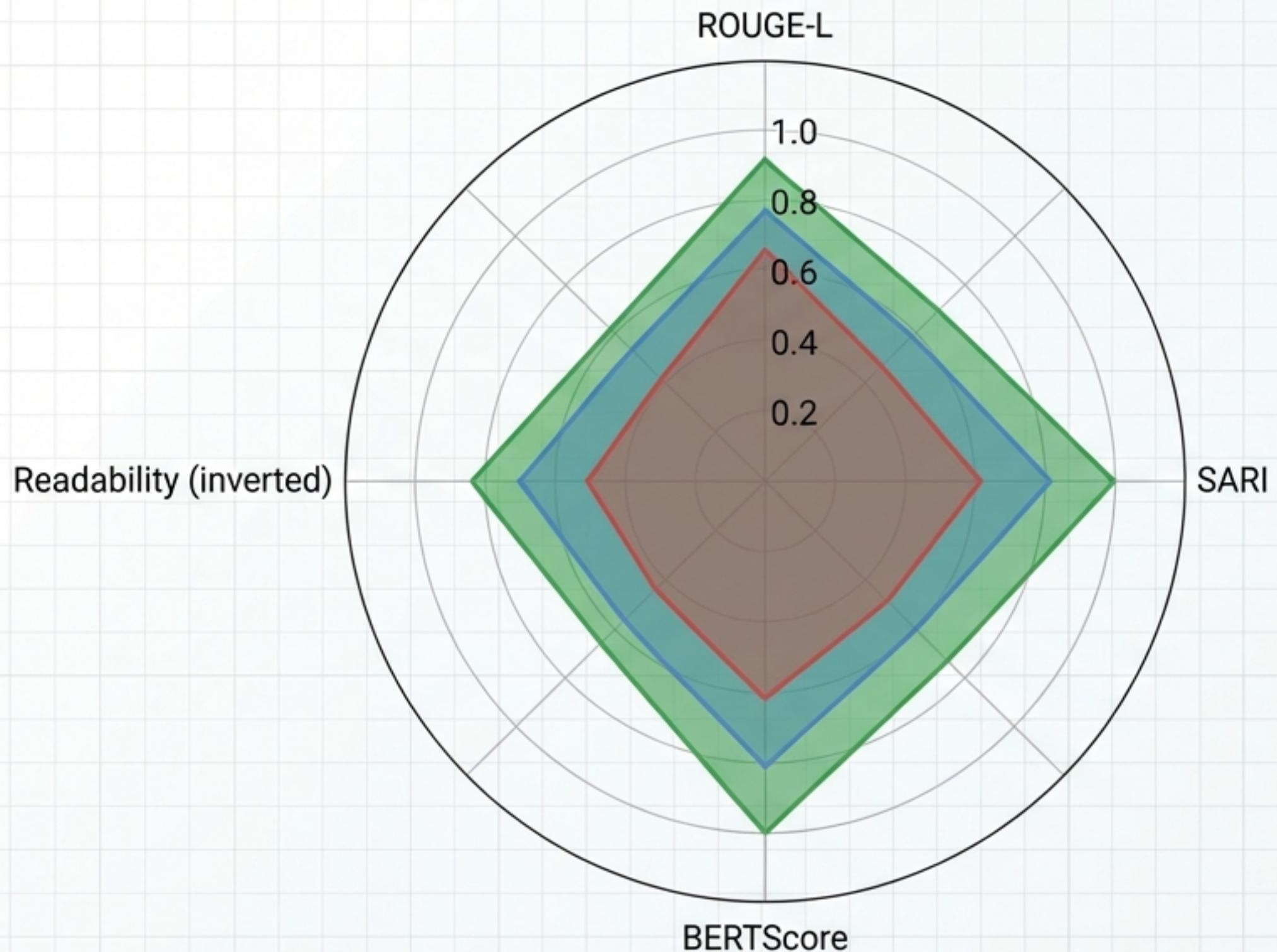
Full Training Configuration



The “Gold” Configuration

- **Rank (r):** 32
- **Alpha (a):** 64
- **Modules:** all_attn (q, k, v, o projections)
- **Method:** rsLoRA (Rank Stabilized LoRA - Crucial for stability)
- **Data:** 7,999 samples
- **Epochs:** 3

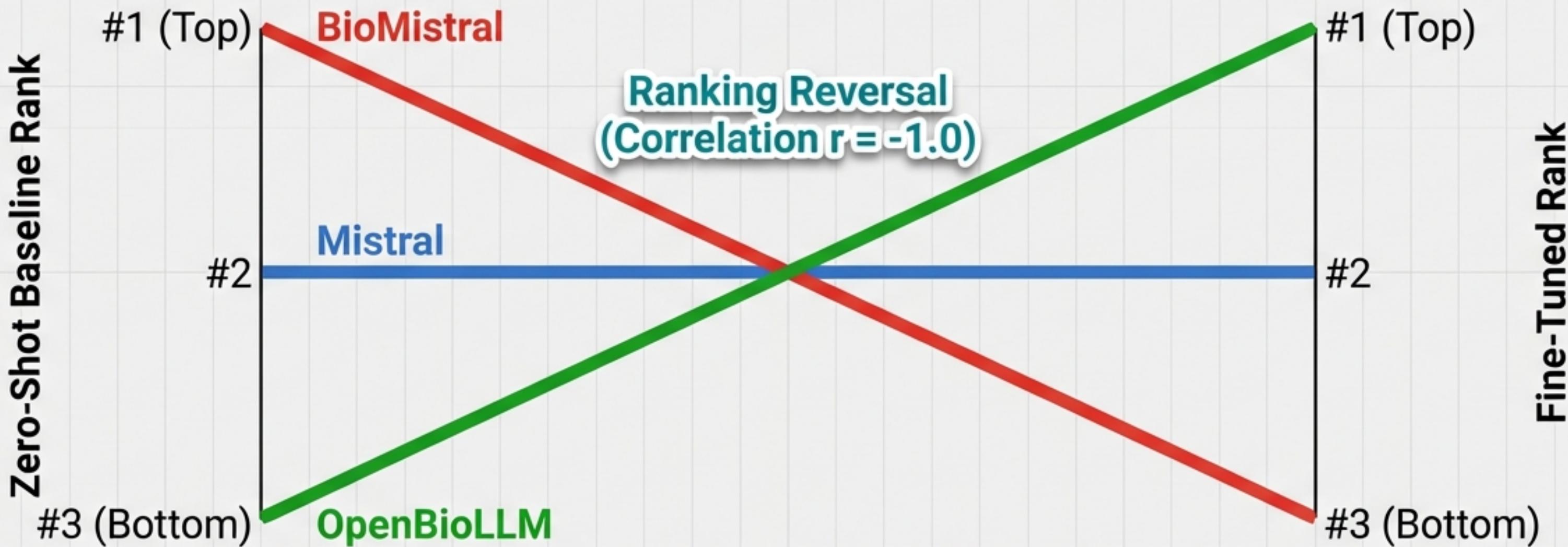
Model Performance Comparison (Post-Training)



Leaderboard (ROUGE-L)

1. OpenBioLLM-8B: 0.675 ROUGE-L
2. Mistral-7B: 0.649 ROUGE-L
3. BioMistral-7B: 0.632 ROUGE-L

KEY FINDING: Ranking Reversal



The ‘Underdog’ had the highest capacity to learn.

Project Results Overview

Improvement

+157% ↑

ROUGE-L increase for
OpenBioLLM.

Readability

14.5 (College)



7.16 (7th Grade)

50% Reduction

Consistency



> 0.94

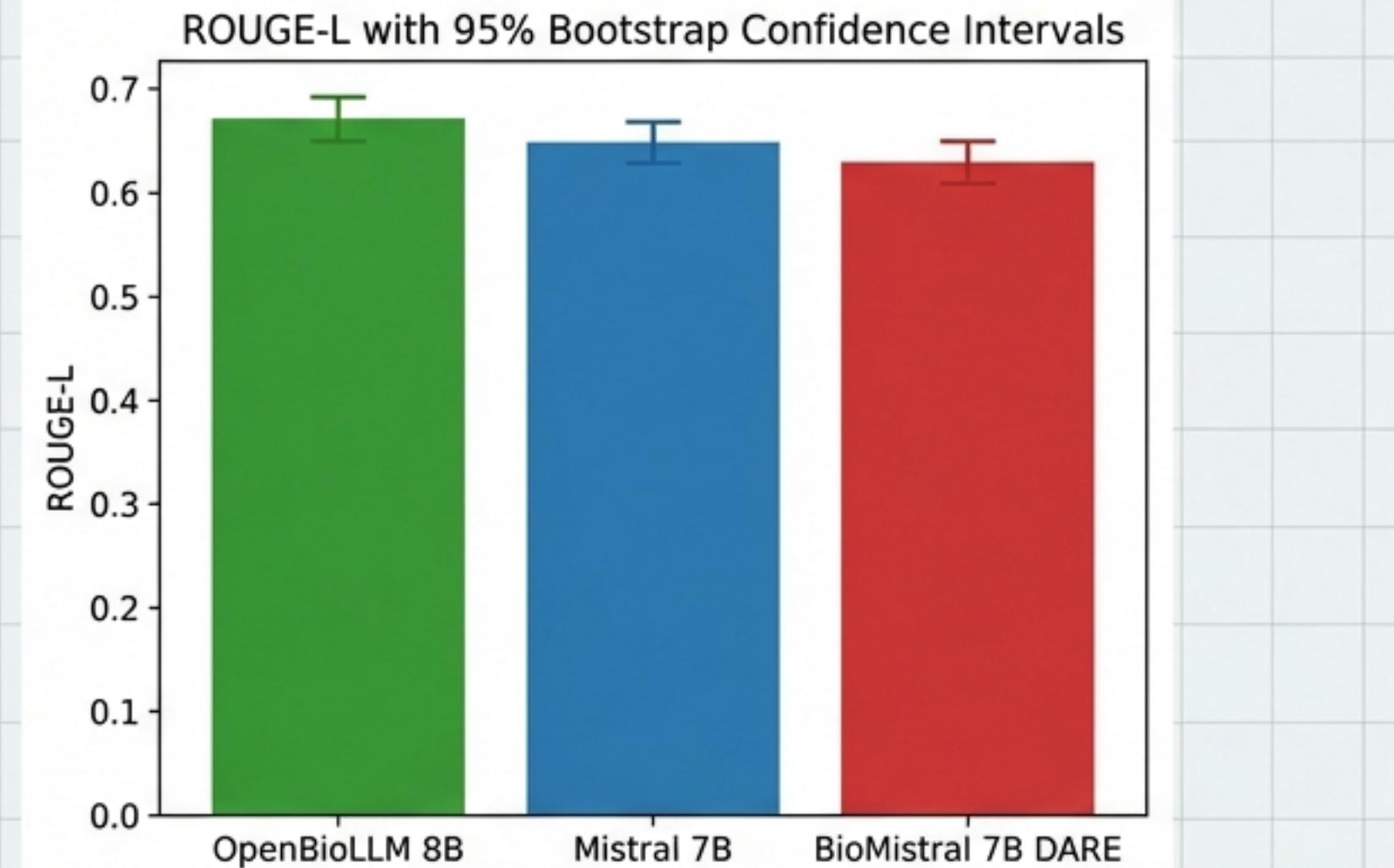
BERTScore across all models.

SARI

74.64 ⏵

Exceeded target of 40.

Statistical Analysis



Effect Size (Cohen's d):

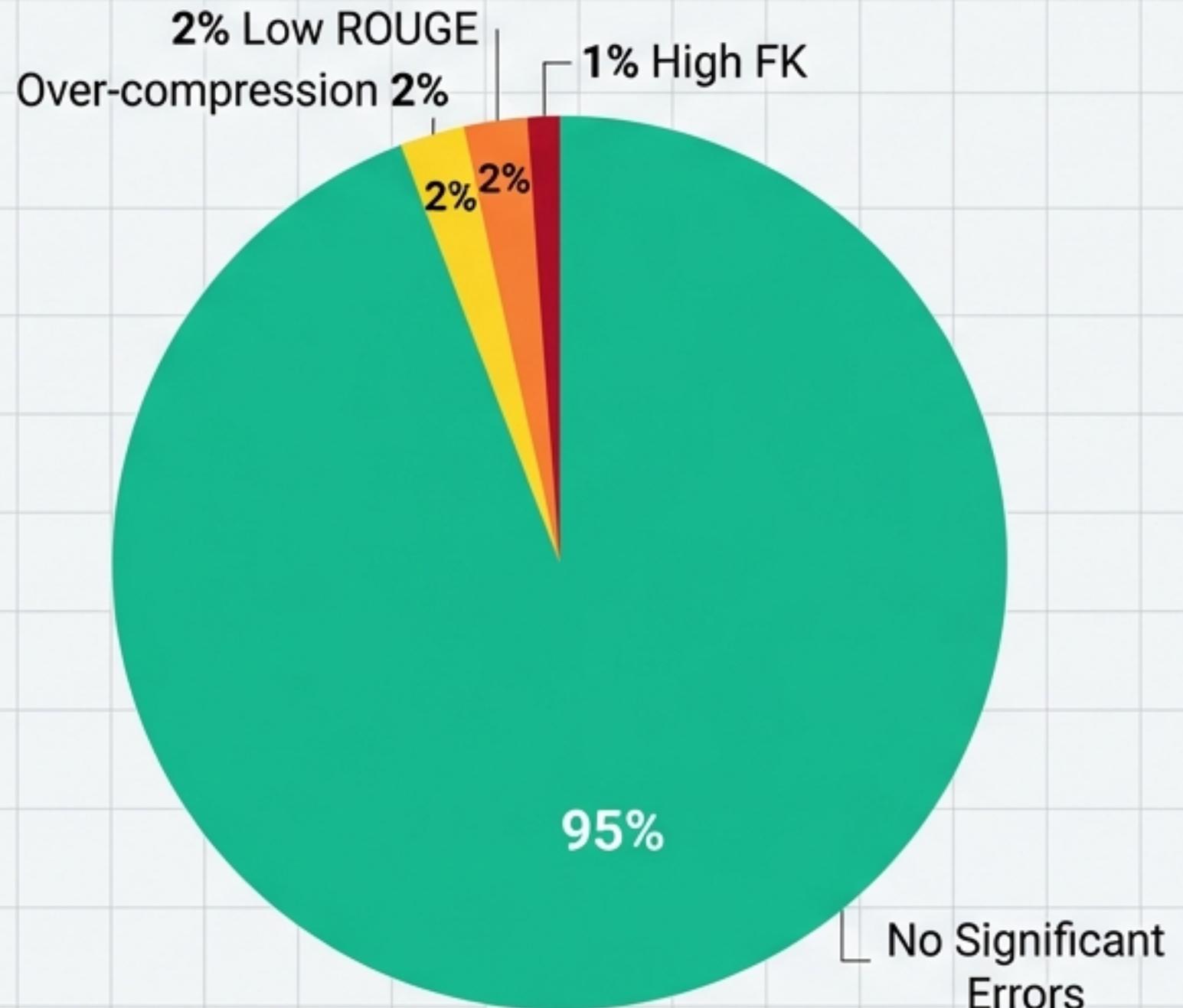
OpenBioLLM vs BioMistral: $d=0.793$
(Medium)

OpenBioLLM vs Mistral: $d=0.475$
(Small)

These effect sizes quantify the magnitude of the differences observed in the confidence interval chart.

**No overlap = Statistically
Significant Differences ($p < 0.001$)**

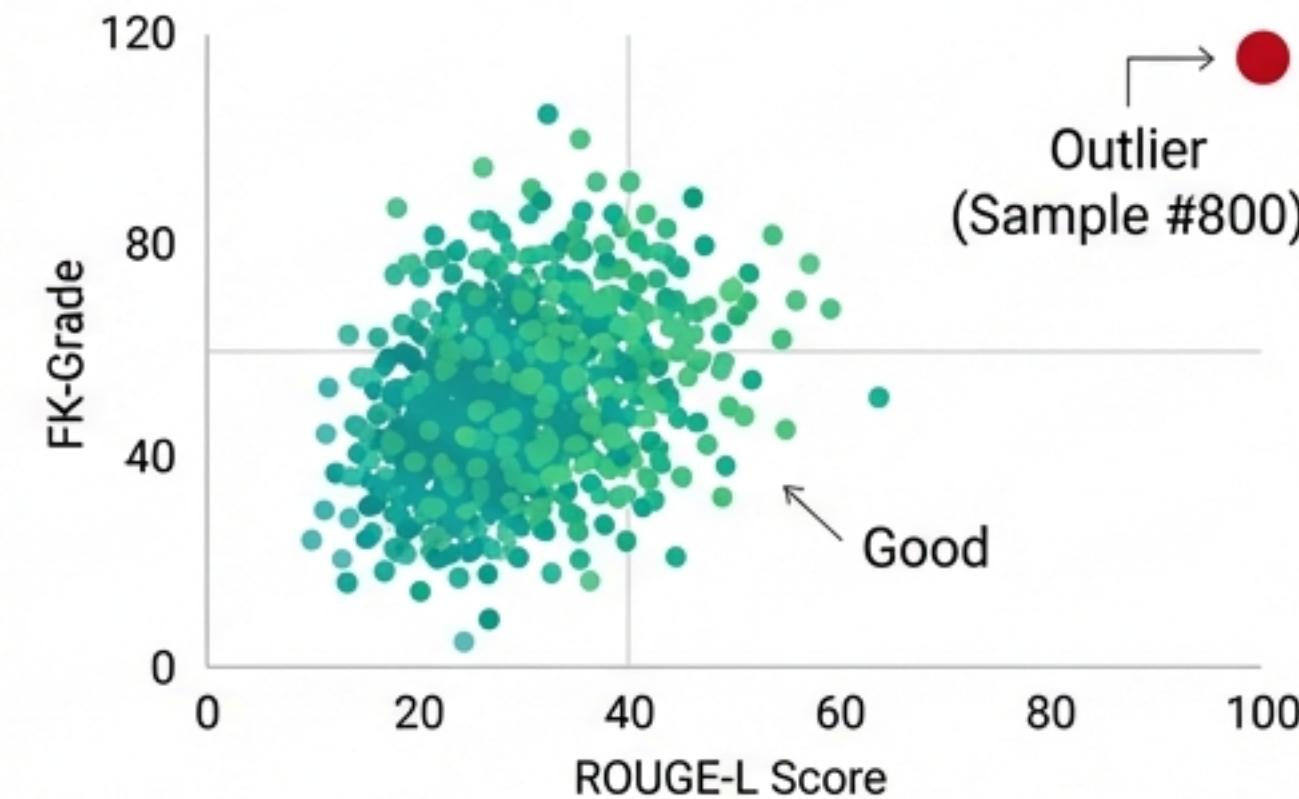
Error Analysis Dashboard



The Outlier

Sample #800: FK Grade = 69.1

Cause: Failed to break up run-on sentences.
Occurred in only 0.1% of data.



Research Questions: Answered

RQ1

Medical pretraining?
No benefit for
Fine-tuning.

RQ2

BioMistral-7B-DARE
best zero-shot

RQ3

LoRA Impact?
+157%
Improvement.

RQ4

Optimal Rank?
 $r=32$.
Contradicts Literature

RQ5

Ranking?
Reversed.

RQ6

Modules?
all_attn optimal.

RQ7

More data = better,
 $4K \approx 97\%$ of $8K$

RQ8

Inverse correlation:
worst → best

RQ9

Parameter cost
justified by quality

RQ10

OpenBioLLM-8B
most consistent
(3/4 wins)

RQ11

~50% FK reduction
(14.5 → 7.0)

RQ12

rsLoRA?
Required for
stability.

Confirmed/Positive Finding

Reversed/Negative Finding

Contradicts Literature

Neutral/Other NotebookLM

Key Takeaways



LoRA Works. ~92% average improvement with only 0.38% trainable parameters.



Domain Value Unlocked. Medical pretraining wasn't visible in Zero-Shot, but dominated after Fine-Tuning.



Readability Bridged. Successfully moved from College Level (14.5) to 7th Grade (7.1).

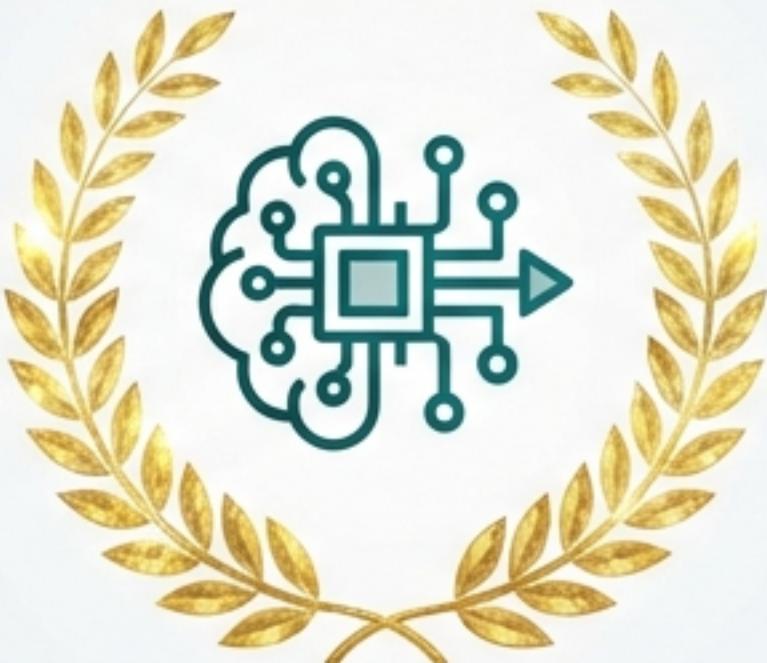
Final Performance Summary

Champion Card

OpenBioLLM-8B

 ROUGE-L:
0.6749
(+157%)

 BERTScore:
0.95
..



 SARI:
74.64
..

 FK-Grade:
7.16
..

The clear winner across all quality metrics.