Model Response Quality Distribution Comparison

Kernel density estimate plots showing distribution of evaluation metrics

ROUGE-1 Score

- Base Model
- Base Mean: 0.175
- QLoRA Model
- QLoRA Mean: 0.197

+12.57

ROUGE-2 Score

- Base Model
- Base Mean: 0.016
- QLoRA Model
- QLoRA Mean: 0.029

+79.48

ROUGE-L Score

- Base Model
- Base Mean: 0.086
- QLoRA Model
- QLoRA Mean: 0.107

+24.00

BLEU Score

- Base Model
- Base Mean: 0.005
- QLoRA Model
- QLoRA Mean: 0.011

+135.36

FinsightAI Evaluation 20250306_121621