ROC Curves for Similarity Metrics 100% -90% 80% 70% True Positive Rate 60% 50% 40% 30% 20% 10% 0% 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% **False Positive Rate**

Metric

- BERTScore (AUC=0.68)
- BLEU (AUC=0.66)
- Classifier (AUC=0.81)
- GPT-3.5-turbo (AUC=0.73)
- GPT-4o-mini (AUC=0.76)— ROUGE-1 (AUC=0.68)
- Random (AUC=0.5)
 - ranaom (1100-0.0)