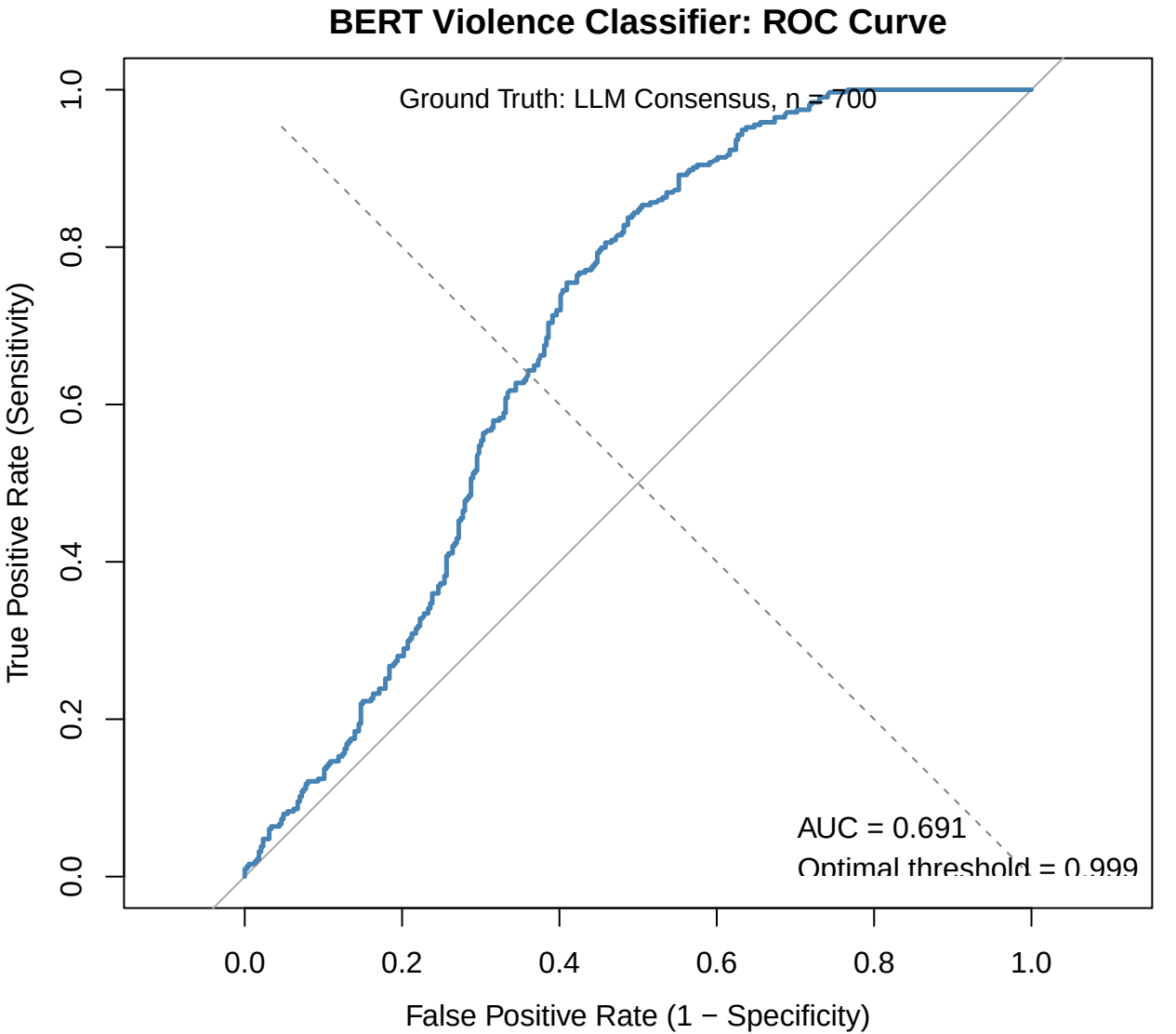
BERT Violence Classifier: ROC Curve
Ground Truth: LLM Consensus, n = 700
AUC = 0.691
Optimal threshold = 0.999
True Positive Rate (Sensitivity)
False Positive Rate (1 − Specificity)
0.0
0.2
0.4
0.6
0.8
1.0