

Table 1: BiLSTM Performance on 13 Clause Categories

Category	Count	EM	F1	Precision	Recall	Answerability	AUC
Insurance	102	0.0%	0.5%	0.5%	0.5%	31.4%	31.4%
Parties	102	0.0%	7.3%	7.3%	7.3%	100.0%	100.0%
Macro Avg	204	0.0%	3.9%	3.9%	3.9%	65.7%	65.7%

Table 2: BiLSTM Uniformed Performance Metrics (13 Categories)

Metric	Macro Avg	Weighted Avg	Std Dev	Range
EM	0.0%	0.0%	0.0%	0.0% – 0.0%
F1	3.9%	3.9%	3.4%	0.5% – 7.3%
Precision	3.9%	3.9%	3.4%	0.5% – 7.3%
Recall	3.9%	3.9%	3.4%	0.5% – 7.3%
Answerability	65.7%	65.7%	34.3%	31.4% – 100.0%
AUC	65.7%	65.7%	34.3%	31.4% – 100.0%