

## Precondition Identification

Model	Precision	Recall	F1
Random	37.34	50.00	42.75
GloVe-GRU	56.25	73.38	63.68
BERT-feature	59.80	81.15	68.84
XLNet	66.69	77.10	71.52
BERT	64.65	81.02	71.91

Table 4: Benchmarking performance of existing models on the precondition identification task. Simply fine-tuning large language models is not enough.