

Metric	M_1	M_2	Averaged	Gold Model	Aggregate Model	Tuned Aggregate Model
Accuracy	0.51	0.53	0.42	0.95	0.71	0.90
Loss	2.79	2.43	1.72	0.16	1.264	0.66