1 Expanded version of Table 14

we have revised Table 14 to include baseline models trained on the proposed dataset and evaluated them across multiple domains, including Restaurants, Wikipedia, Medicine, Essays, E-commerce, Applied Statistics, and M4. An expanded version of Table 14—reporting Accuracy and F1-scores for the binary classification task (Task A) on these domains for all baselines from Table 13—is provided in the accompanying PDF at the following anonymous link.

Model	**Restaurant**		**Wikipedia**		**Medicine**		**Essay**		**E-commerce**		**Applied Stats**		**M4**	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
GTLR	0.7400	0.590	0.7700	0.520	0.7200	0.570	0.6700	0.630	0.6800	0.580	0.7000	0.560	0.5100	0.540
OpenAI-detector	0.7600	0.610	0.7800	0.620	0.7400	0.590	0.7200	0.580	0.7500	0.630	0.7100	0.600	0.5300	0.520
DetectGPT	0.7900	0.740	0.8100	0.770	0.8000	0.760	0.7300	0.790	0.7900	0.780	0.7200	0.700	0.5800	0.610
Radar	0.7200	0.640	0.7600	0.680	0.7700	0.700	0.7000	0.670	0.7300	0.720	0.7000	0.660	0.5500	0.590
RoBERTa	0.5037	0.400	0.7481	0.500	0.7902	0.620	0.6836	0.650	0.7065	0.600	0.7056	0.550	0.5511	0.570
BART	0.6954	0.500	0.7083	0.410	0.6826	0.480	0.8003	0.640	0.7058	0.740	0.7092	0.670	0.5962	0.560
Ours	**0.8000**	**0.848**	**0.8100**	**0.890**	**0.8200**	**0.883**	**0.7500**	**0.886**	**0.8500**	**0.885**	**0.7500**	**0.791**	**0.6700**	**0.770**

Table 1: Accuracy and F1 score of various detectors across different domains.