Table A1. Evaluation of downstream STM accuracy using 6PLMs, $L = 1$. Experimental settings are the same as in *Table 1*. Method "DP-SGD+Gen" first fine-tunes the PLM under DP protection using DP-SDG and then use the fine-tuned model for synthetic data generation. **Best** and second best results are marked.

		Privacy	$\lvert\mathcal{B}\rvert$	$\lvert\mathcal{D}\rvert$	IMDb	Yelp Category	Yelp Rating	Openreview Area	Openreview Rating	Banking
OnlyPrivate		$\epsilon = \infty$	100	-	50.00	5.69	35.57	6.56	22.20	13.75
FuseGen		Absolutely Private	-	6,000	89.07	63.38	57.96	24.70	34.57	78.75
DP-SGD+Gen	GPT-2	$\epsilon = 4.0$	100	6,000	87.44	47.45	50.04	33.20	31.25	74.88
	Llama-2	$\epsilon = 4.0$	100	6,000	84.63	62.14	49.95	28.23	28.45	79.75
	Vicuna	$\epsilon = 4.0$	100	6,000	84.93	62.99	57.46	31.17	23.48	78.75
	OPT	$\epsilon = 4.0$	100	6,000	81.47	62.61	55.68	34.57	22.00	75.75
	ChatGLM3	$\epsilon = 4.0$	100	6,000	83.17	52.99	45.79	34.60	33.99	84.38
	Flan-T5	$\epsilon = 4.0$	100	6,000	89.14	58.59	60.85	33.52	35.35	78.13
Aug-PE	GPT-2	$\epsilon = 4.0$	100	6,000	85.38	62.33	45.28	31.45	24.12	75.63
	Llama-2	$\epsilon = 4.0$	100	6,000	85.77	60.18	47.42	32.67	34.78	84.63
	Vicuna	$\epsilon = 4.0$	100	6,000	82.76	63.28	54.42	32.27	30.66	86.75
	OPT	$\epsilon = 4.0$	100	6,000	83.86	62.71	50.81	34.64	25.30	79.25
	ChatGLM3	$\epsilon = 4.0$	100	6,000	85.82	55.06	55.17	33.81	32.49	88.50
	Flan-T5	$\epsilon = 4.0$	100	6,000	89.00	62.06	58.69	34.54	35.42	81.25
WASP (Ours)		$\epsilon = 4.0$	100	6,000	**89.52**	**63.91**	**61.21**	**34.99**	**37.10**	**88.75**

Table F1. Comparison of downstream STM accuracy under w/ and w/o Contrastive In-context Learning and Private Data Assisted PLM Importance Weighting setting using 6 open-source PLMs, $L = 1$. DP setting is same as that used for *Table 1*. "w/o both" indicates that both techniques are removed with only Top-Q Voting with $Q = 8$ remains.

	w/o both	w/o PLM Contrastive Prompting	w/o PLM Importance Weighting	WASP (Ours)
IMDb	89.05	89.21	89.17	89.52
Yelp-Rating	58.72	59.65	58.94	61.21
Openreview-Rating	35.45	36.18	35.53	37.10