

1 Baseline (un-fine-tuned) model performance

2 Stance Detection

Table 1: Classification Report for GPT-4 Few-shot as target labels: Un-finetuned performance: weighted F1 score for Stance classification using SLMs on USDC test set.

Class	Precision	Recall	F1-Score	Support
Somewhat Against	0.26	0.67	0.38	400
Somewhat In Favor	0.45	0.21	0.28	624
Stance Not Inferrable	0.35	0.11	0.16	454
Strongly Against	0.25	0.38	0.30	261
Strongly In Favor	0.13	0.02	0.03	128
Accuracy		0.29		1867
Macro avg	0.29	0.28	0.23	1867
Weighted avg	0.33	0.29	0.26	1867

Table 2: Classification Report for Mistral Large few-shot as target labels: Un-finetuned performance: weighted F1 score for Stance classification using SLMs on USDC test set.

Class	Precision	Recall	F1-Score	Support
Somewhat Against	0.20	0.69	0.31	316
Somewhat In Favor	0.39	0.24	0.30	458
Stance Not Inferrable	0.41	0.08	0.14	567
Strongly Against	0.29	0.32	0.30	336
Strongly In Favor	0.31	0.02	0.04	190
Accuracy		0.26		1867
Macro avg	0.32	0.27	0.22	1867
Weighted avg	0.34	0.26	0.23	1867

Table 3: Classification Report for Majority Voting as target labels: Un-finetuned performance: weighted F1 score for Stance classification using SLMs on USDC test set.

Class	Precision	Recall	F1-Score	Support
Somewhat Against	0.30	0.71	0.42	443
Somewhat In Favor	0.41	0.20	0.27	625
Stance Not Inferrable	0.34	0.09	0.14	452
Strongly Against	0.26	0.39	0.31	256
Strongly In Favor	0.19	0.03	0.06	91
Accuracy		0.31		1867
Macro avg	0.30	0.28	0.24	1867
Weighted avg	0.34	0.31	0.27	1867

3 Dogmatism Identification

Table 4: Classification Report for Majority Voting as target labels: Un-finetuned performance: weighted F1 score for Dogmatism classification using SLMs on USDC test set.

Class	Precision	Recall	F1-Score	Support
Deeply Rooted	0.17	0.54	0.26	28
Firm but Open	0.50	0.25	0.34	131
Flexible	0.00	0.00	0.00	14
Open to Dialogue	0.48	0.55	0.51	134
Accuracy		0.40		307
Macro avg	0.29	0.33	0.28	307
Weighted avg	0.44	0.40	0.39	307