Table 1: Evaluation Metrics for DrawBench Models

Model	ALOHA	CIDEr	CLIP	MHalDetection	BLEU-1	BLEU-4	ROUGE-L_F1	BERTScore_F1	METEOR	SPICE	TIFA
Flux-Dev	0.1223	0.03788	0.3147	0.8061	0.0827	0.01130	0.1426	0.8564	0.1817	0.1308	0.8181
SD2	0.06721	0.02940	0.3219	0.9019	0.06616	0.00709	0.1149	0.8527	0.1424	0.09355	0.8181
SDXL	0.08111	0.02496	0.3206	0.8912	0.06553	0.00731	0.1172	0.8517	0.1526	0.1034	0.8181
Mean	0.09088	0.03008	0.3191	0.8664	0.0715	0.00857	0.1249	0.8536	0.1589	0.1099	0.8181
Std.Dev	0.02896	0.00665	0.0038	0.0477	0.0091	0.00212	0.0145	0.0023	0.0160	0.0186	0.0000