

Table 1: Evaluation Metrics for Models on MSCOCO

Model	ALOHA	CIDEr	CLIP	MHalDetect	BLEU-1	BLEU-4	ROUGE-L_F1	ROUGE-L_Recall	BERTScore_F1	BERTScore_Recall	METEOR	SPICE	TIFA
Flux-Dev	0.1693	0.0000485	0.3069	0.9572	0.0460	0.00543	0.1011	0.2809	0.8311	0.8648	0.1357	0.0421	0.87385
SD2	0.1351	0.0025563	0.3112	0.9659	0.0625	0.00894	0.1361	0.3985	0.8445	0.8819	0.1842	0.0861	0.87385
SDXL	0.0513	0.00000894	0.2767	0.9856	0.0683	0.00944	0.1489	0.4351	0.8496	0.8885	0.1975	0.0967	0.87385
Mean	0.1186	0.0008719	0.2983	0.9696	0.0583	0.00727	0.1287	0.3715	0.8411	0.8784	0.1725	0.0749	0.87385
Std.Dev	0.0607	0.0013060	0.0172	0.0140	0.0113	0.00193	0.0206	0.0672	0.0093	0.0103	0.0330	0.0277	0.0000