

Attacked Models

Cross Entropy of P(Target | Prompt, Image)

'prism-siglip+7b'

'prism-clip+7b'

eval_model_str

prism-reproduction-llava-v15+7b

Evaluated Model

prism-reproduction-llava-v15+7b

Gradient Step

Gradient Step

