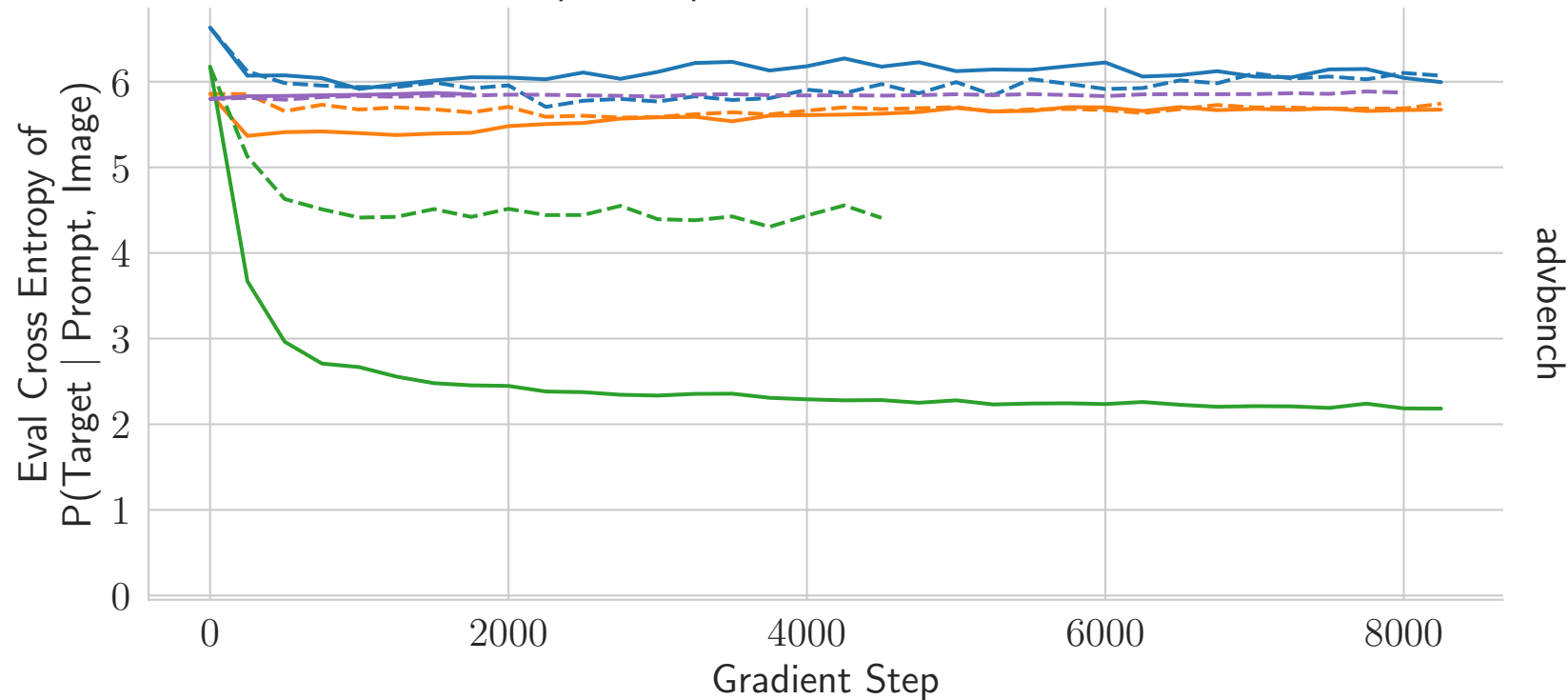


Attacked Model(s)

'prism-reproduction-llava-v15+7b'



Evaluated Model

- model_to_eval
- 'prism-clip+7b'
- 'prism-clip-controlled+7b'
- 'prism-reproduction-llava-v15+7b'
- 'prism-siglip+7b'
- 'prism-siglip-controlled+7b'
- attack_dataset
- advbench
- rylan_anthropic_hhh