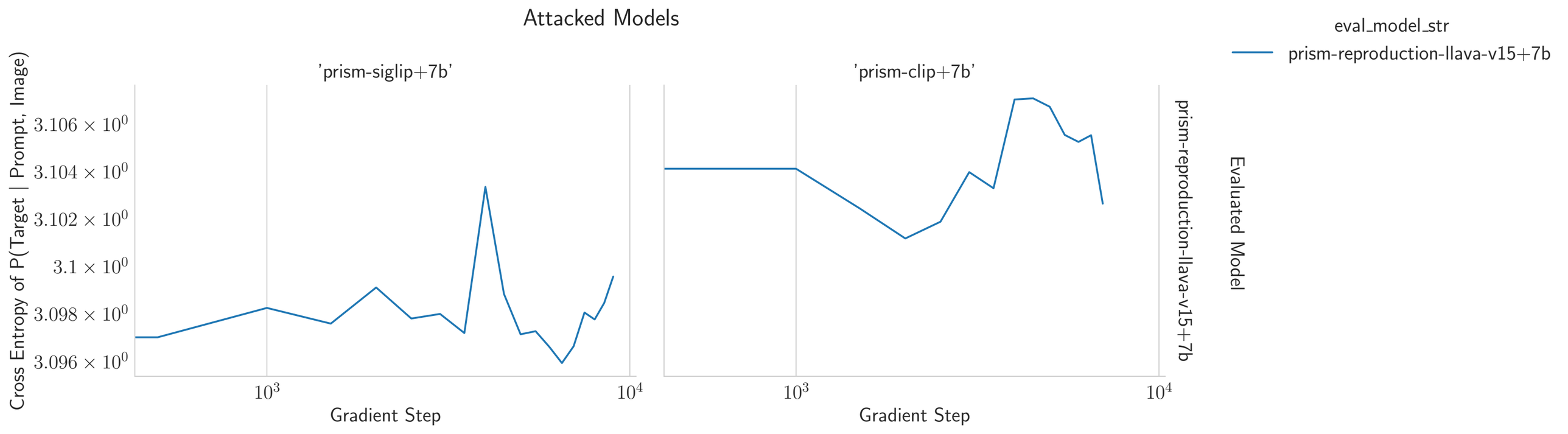
Attacked Models
'prism-siglip+7b'
'prism-clip+7b'
Cross Entropy of P(Target | Prompt, Image)
3.106×10^0
3.104×10^0
3.102×10^0
3.1×10^0
3.098×10^0
3.096×10^0
10^3
10^4
Gradient Step
Evaluated Model
prism-reproduction-llava-v15+7b
eval_model_str
prism-reproduction-llava-v15+7b