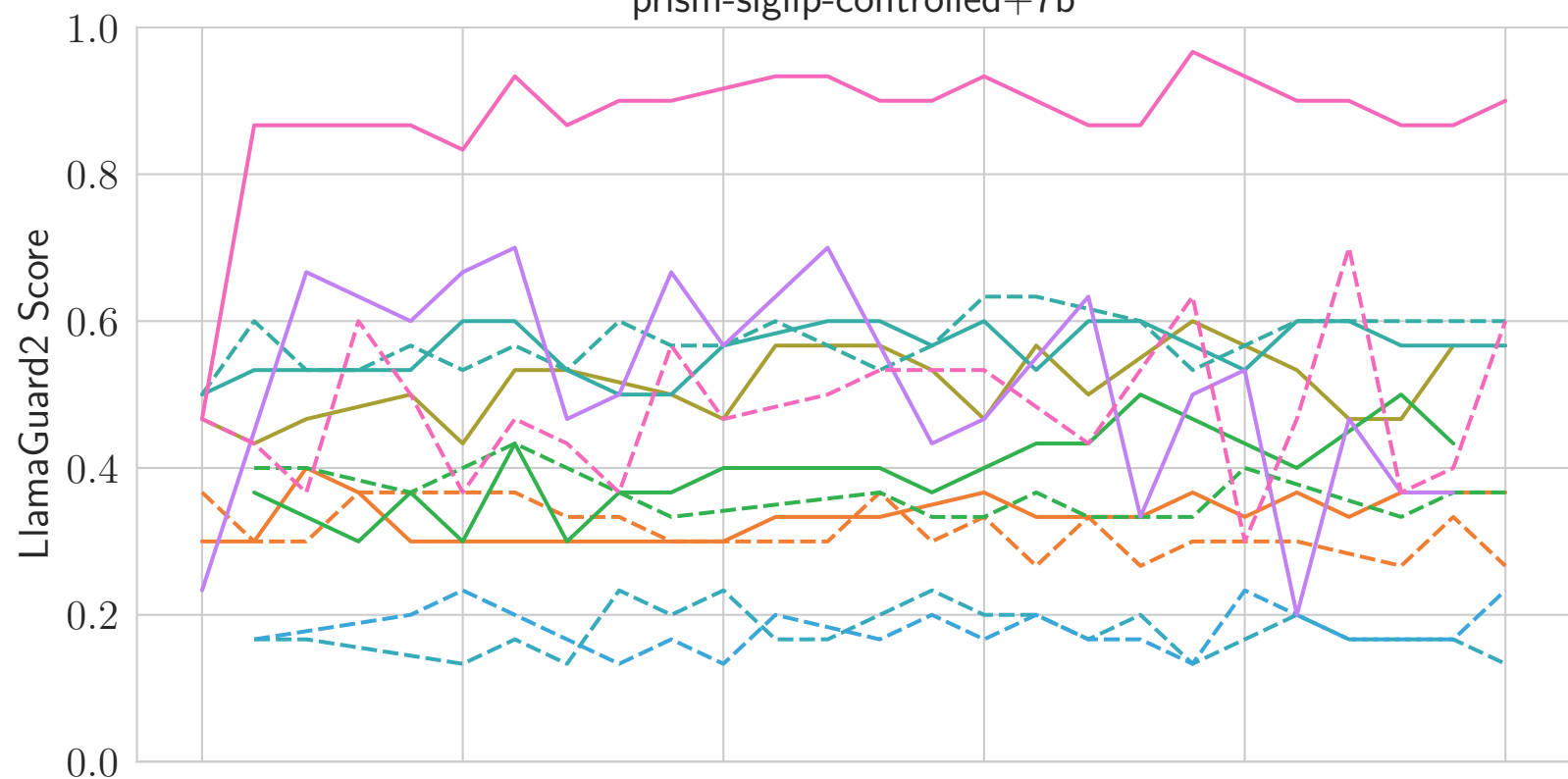
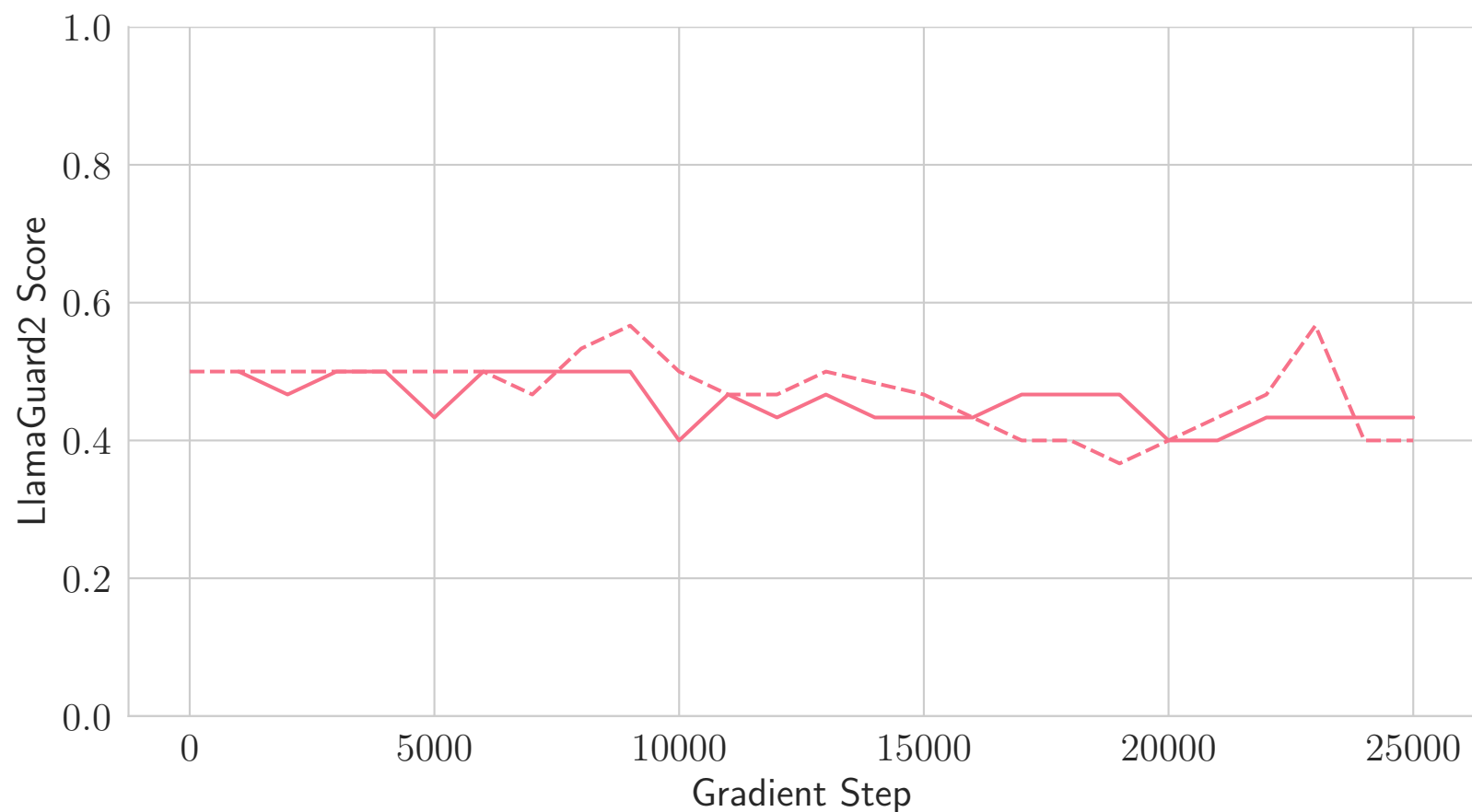


# Attacked Model(s)

'prism-siglip-controlled+7b'



rylan-anthropic-hhh



advbench