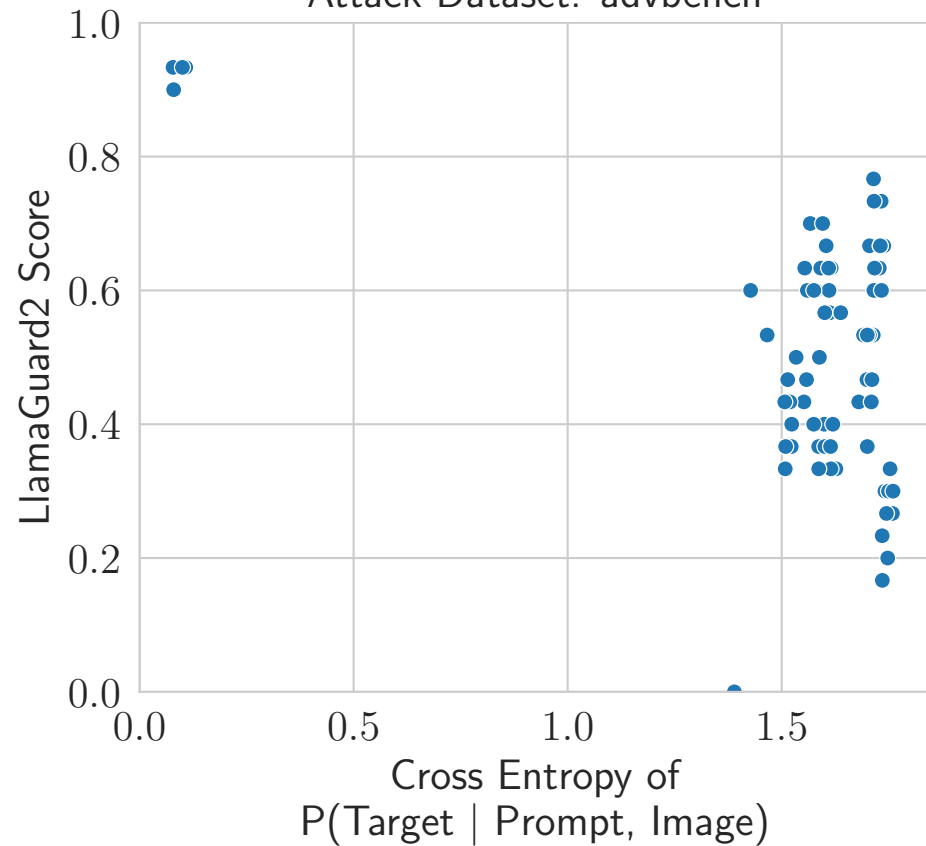


Attack Dataset: advbench



Attack Dataset: rylan\_anthropic\_hhh

