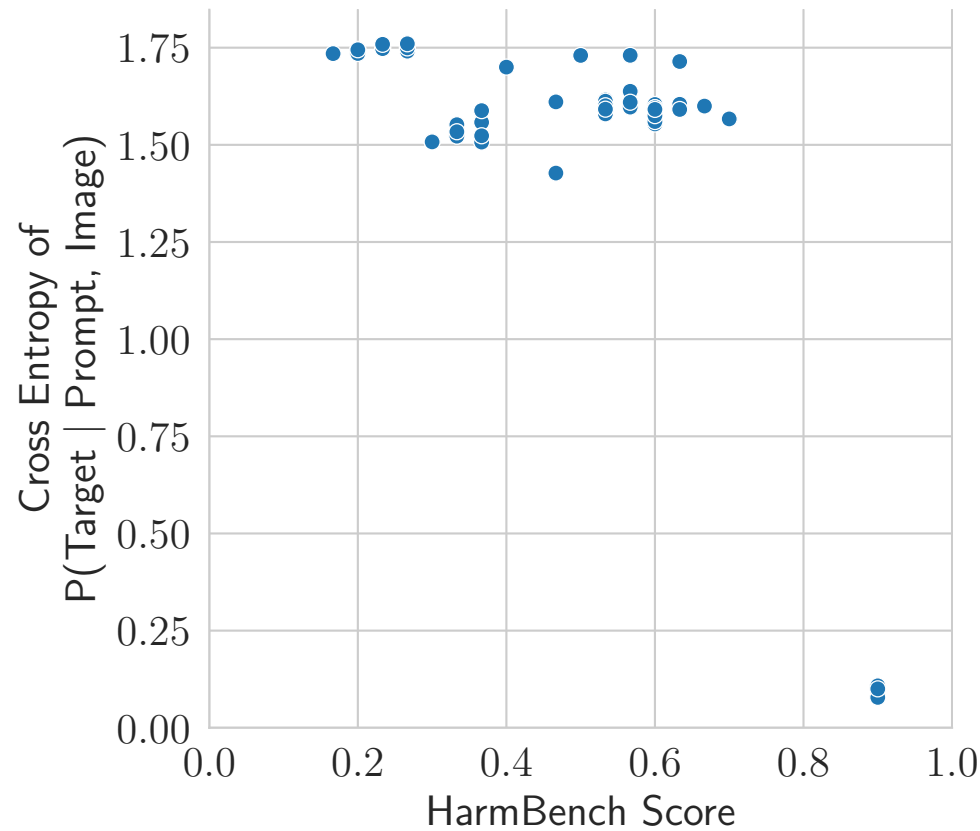


Attack Dataset: advbench



Attack Dataset: rylan_anthropic_hhh

