

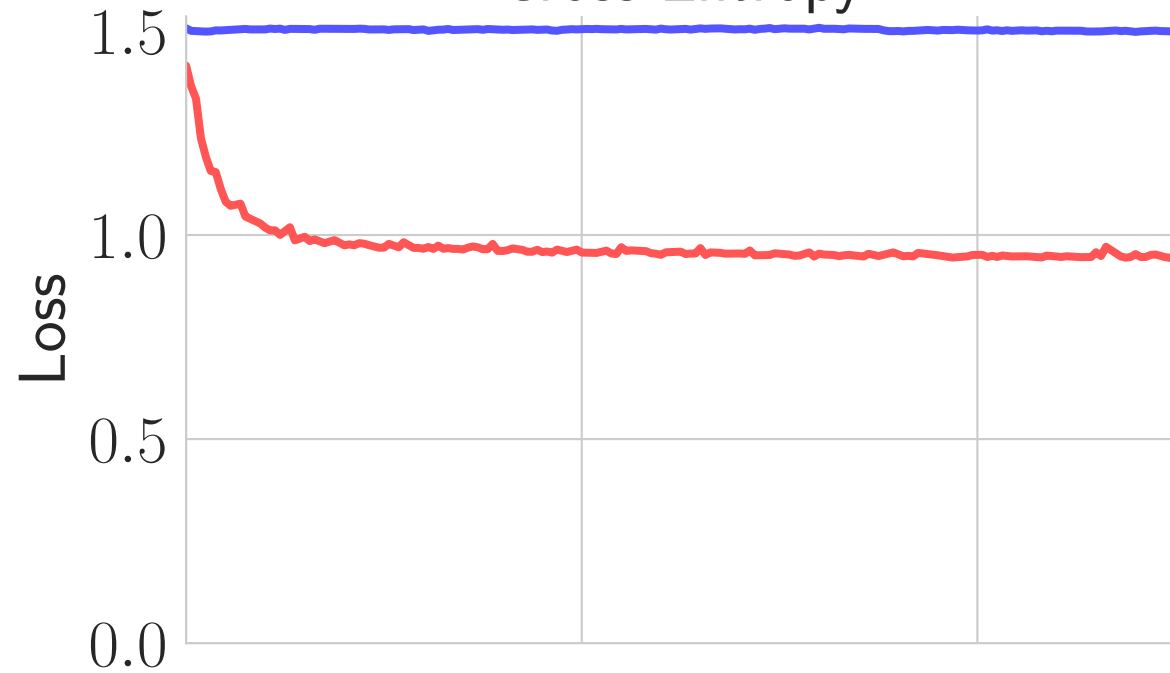
Transfer Between VLM Trained for 1 Stage vs 2 Stages

VLM Training Stages

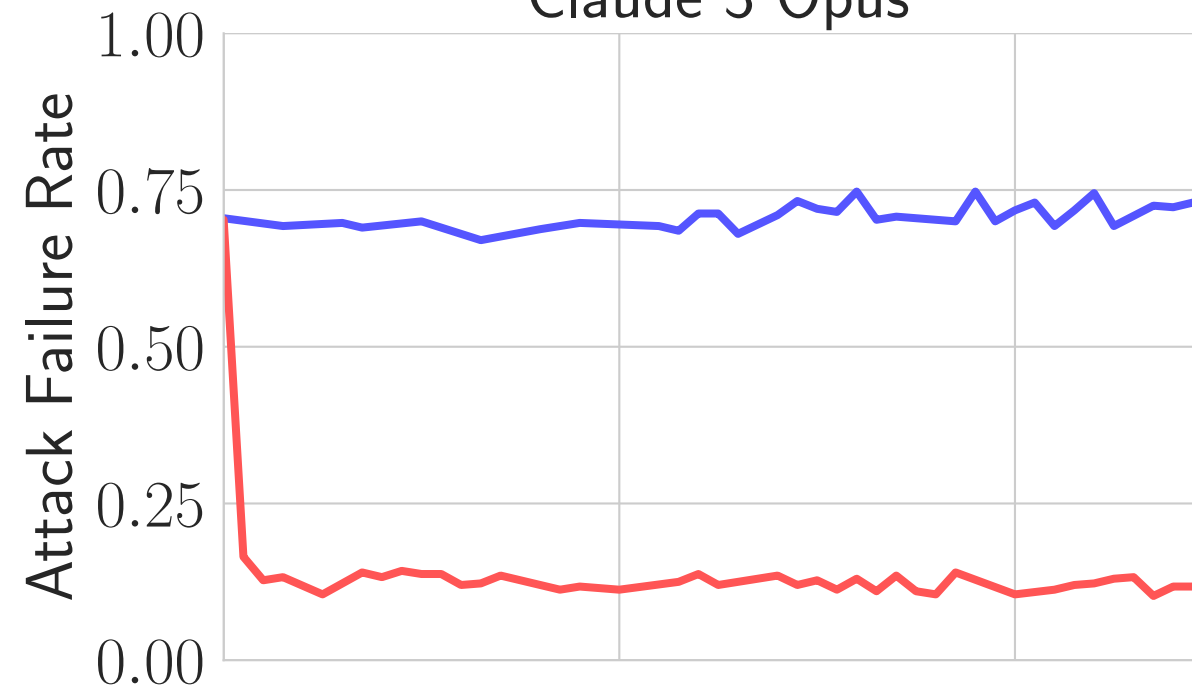
Two-Stage

One-Stage

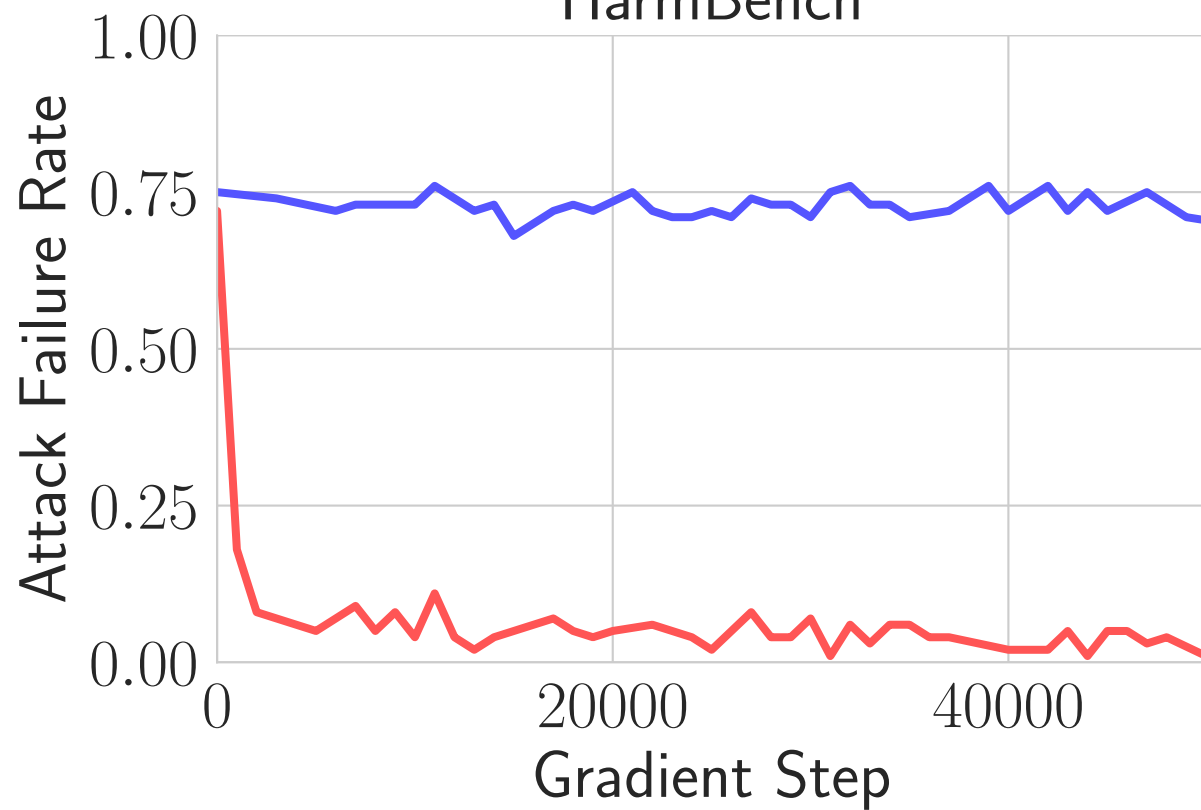
Cross Entropy



Claude 3 Opus



HarmBench



LlamaGuard2

