

HarmBench LLM Judge - Most Likely Token Probability by Temperature

