

Histogram of Perplexity vs Average Harmfulness Evaluation
Risk Area: Malicious Uses

