

Histogram of Perplexity vs Average Harmfulness Evaluation  
Risk Area: Misinformation Harms

