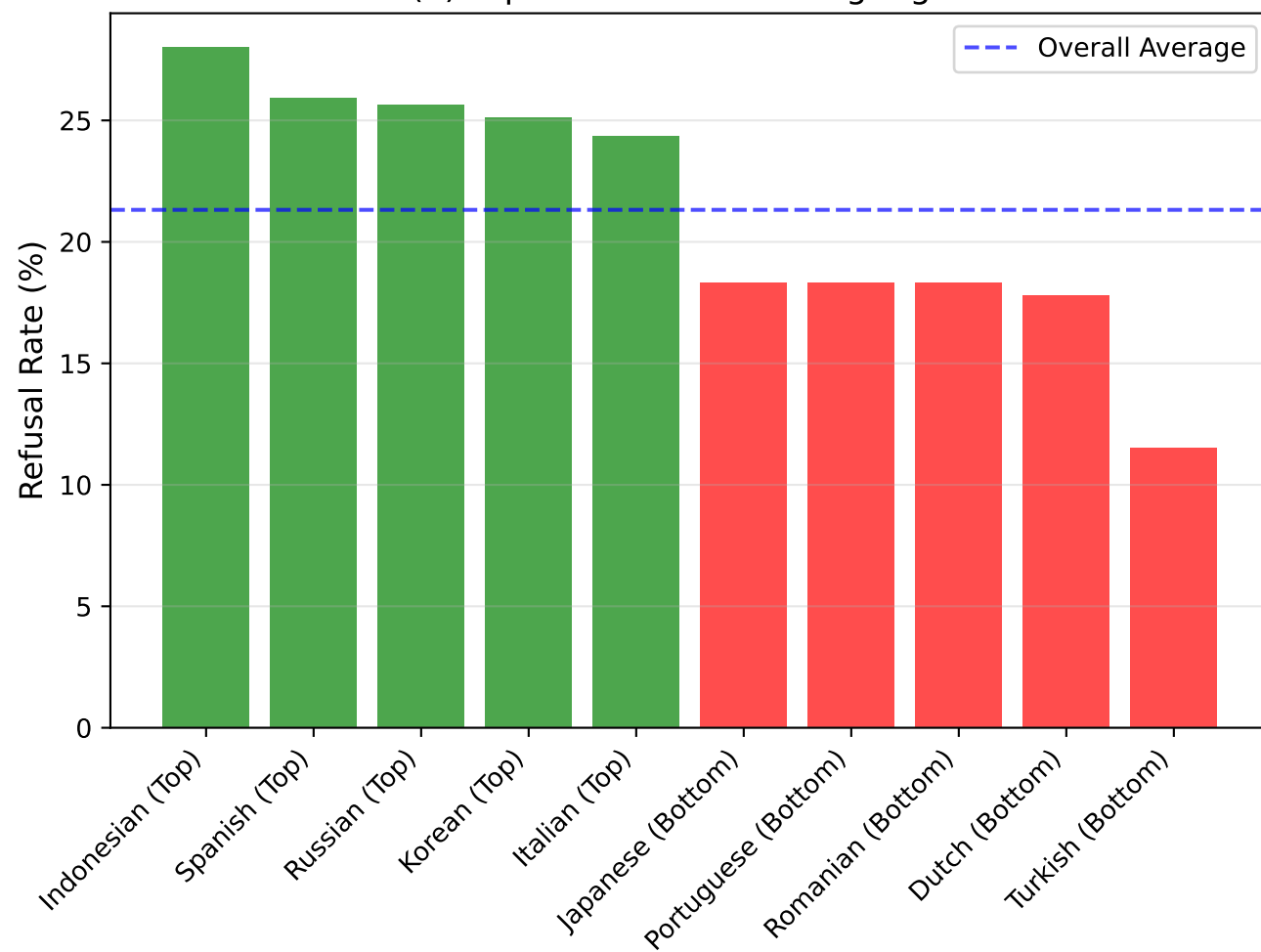
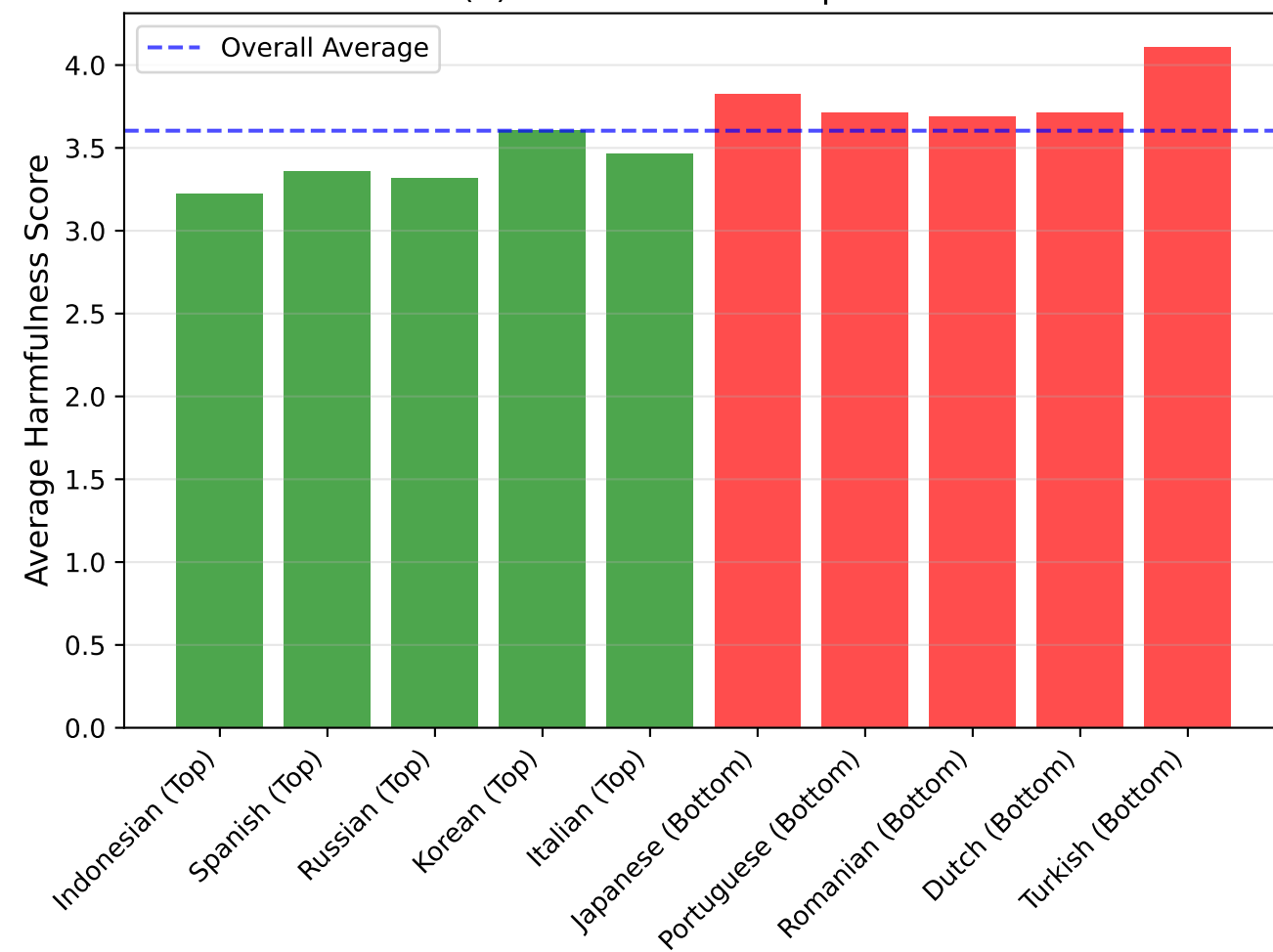


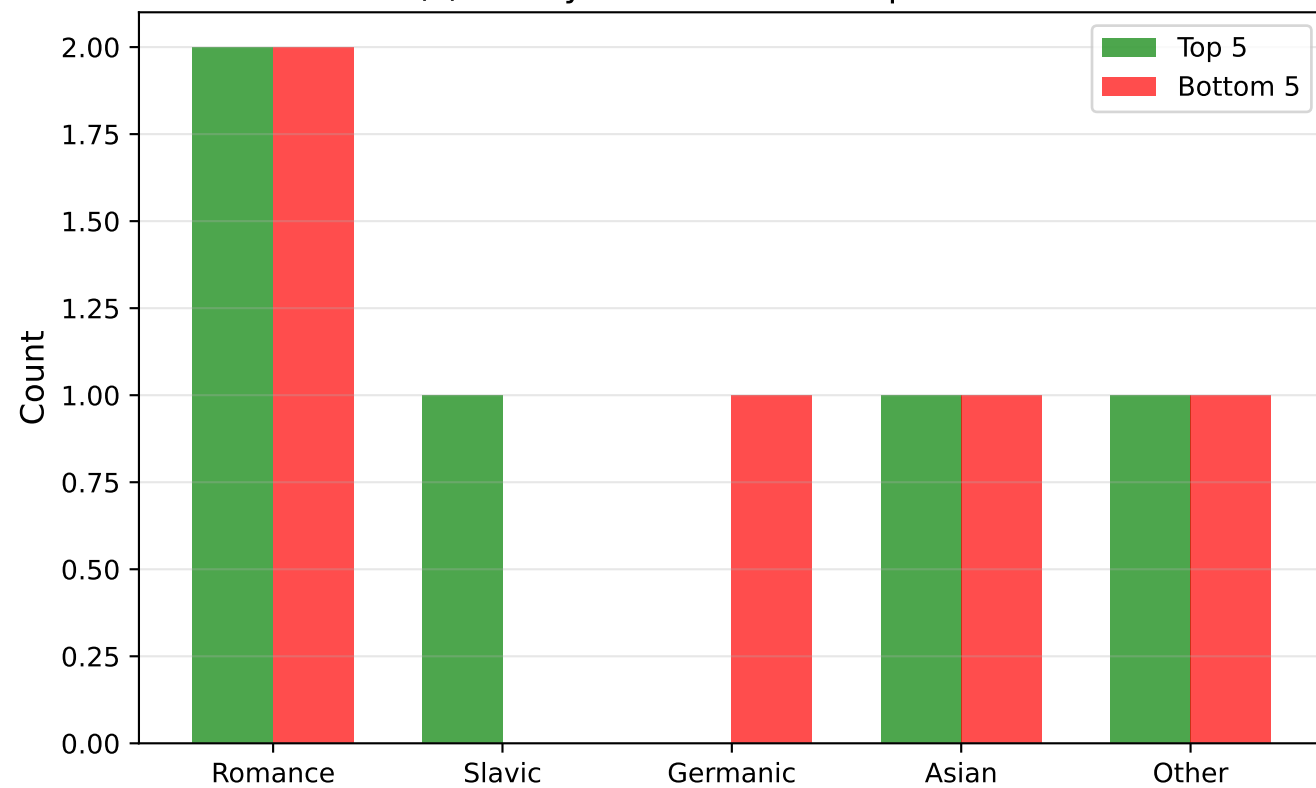
(a) Top 5 vs Bottom 5 Languages



(b) Harmfulness Comparison



(c) Family Distribution in Top/Bottom



## Performance Summary:

## Top 5 Languages (Safest):

- Average Refusal: 25.8%
- Average Score: 3.40
- Range: 24.3% - 28.0%

## Bottom 5 Languages (Most Vulnerable):

- Average Refusal: 16.9%
- Average Score: 3.81
- Range: 11.5% - 18.3%

## Gap Analysis:

- Refusal Rate Gap: 9.0%
- Score Gap: 0.42

## Statistical Significance:

- p-value < 0.001 (highly significant)