Safety Metrics by Language - aya-expanse Harmful Trend **Refusal Trend** 70 Harmful Responses Refusal Rate 60 50 Rate (%) 30 20 10 0 ·X. i ox 54 er 公 1/2 40 Language