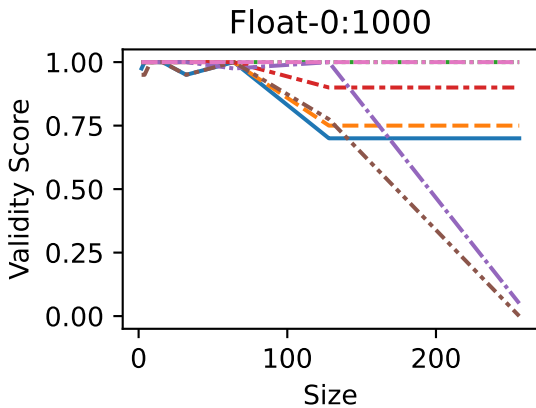
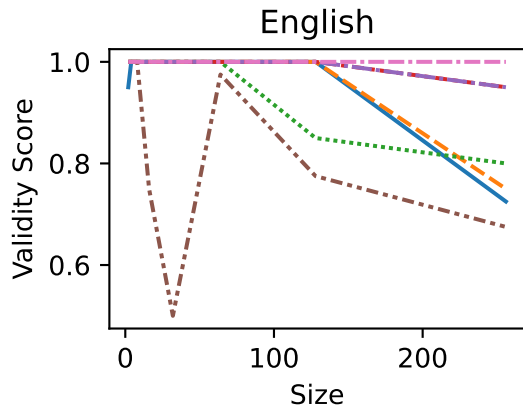
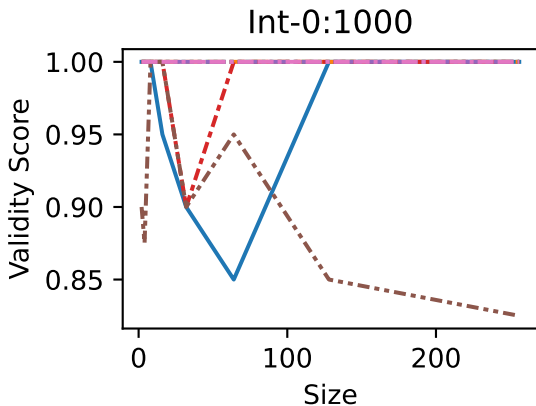


# Validity Score



- Claude-3.5-Haiku
- Claude-3.5-Sonnet
- GPT-4o
- GPT-4o-mini
- Llama-3.1
- DeepSeek-r1
- o3-mini