Israel-USA — No Refusal — Hebrew

Model	Average WD (topic average)
gpt-4o-mini (hebrew)	0.348
Qwen2.5-1.5B-Instruct (hebrew)	0.162
c4ai-command-r7b (hebrew)	0.158
Qwen2.5-3B-Instruct (hebrew)	0.143
zephyr-7b-beta (hebrew)	0.138
Phi-3-mini-Instruct (hebrew)	0.138
Qwen2.5-7B-Instruct (hebrew)	0.136
Gemma3-1b-Instruct (hebrew)	0.134
mistral-7b-v0.3 (hebrew)	0.134
llama-3-8b-Instruct (hebrew)	0.134
Phi-3-medium-Instruct (hebrew)	0.129

Language
- english
- hebrew