Summary — No Refusal — All

Model	Average WD (file-level averages)
gpt-4o-mini (hebrew)	0.347
gpt-4o-mini (english)	0.338
Gemma3-1b-Instruct (english)	0.200
c4ai-command-r7b (english)	0.183
zephyr-7b-beta (english)	0.181
Gemma3-1b-Instruct (hebrew)	0.179
c4ai-command-r7b (hebrew)	0.179
Qwen2.5-3B-Instruct (hebrew)	0.177
Qwen2.5-1.5B-Instruct (hebrew)	0.177
Qwen2.5-1.5B-Instruct (english)	0.168
llama-3-8b-Instruct (english)	0.166
mistral-7b-v0.3 (hebrew)	0.165
zephyr-7b-beta (hebrew)	0.165
Qwen2.5-7B-Instruct (hebrew)	0.164
llama-3-8b-Instruct (hebrew)	0.164
Phi-3-medium-Instruct (hebrew)	0.161
Phi-3-mini-Instruct (hebrew)	0.160
Phi-3-medium-Instruct (english)	0.158
Qwen2.5-3B-Instruct (english)	0.154
Phi-3-mini-Instruct (english)	0.151
mistral-7b-v0.3 (english)	0.146
Qwen2.5-7B-Instruct (english)	0.132

Language: english, hebrew