International Relations \ Foreign policy gpt-4o-mini (hebrew) Qwen2.5-1.5B-Instruct (hebrew) Qwen2.5-3B-Instruct (hebrew) Qwen2.5-7B-Instruct (hebrew) zephyr-7b-beta (hebrew) Gemma3-1b-Instruct (hebrew) Phi-3-mini-Instruct (hebrew) c4ai-command-r7b (hebrew) Phi-3-medium-Instruct (hebrew) llama-3-8b-Instruct (hebrew) mistral-7b-v0.3 (hebrew) 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 Average WD Israel-Palestine gpt-4o-mini (hebrew) 0.198 Gemma3-1b-Instruct (hebrew) Qwen2.5-1.5B-Instruct (hebrew) mistral-7b-v0.3 (hebrew) Phi-3-medium-Instruct (hebrew) llama-3-8b-Instruct (hebrew) c4ai-command-r7b (hebrew) Phi-3-mini-Instruct (hebrew) Qwen2.5-3B-Instruct (hebrew) zephyr-7b-beta (hebrew) 0.152 Qwen2.5-7B-Instruct (hebrew) 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 Average WD Israel-USA gpt-4o-mini (hebrew) Qwen2.5-1.5B-Instruct (hebrew) c4ai-command-r7b (hebrew) Gemma3-1b-Instruct (hebrew) Qwen2.5-3B-Instruct (hebrew) llama-3-8b-Instruct (hebrew) mistral-7b-v0.3 (hebrew) Phi-3-mini-Instruct (hebrew) Phi-3-medium-Instruct (hebrew) Qwen2.5-7B-Instruct (hebrew) zephyr-7b-beta (hebrew) 0.05 0.35 0.40 0.10 0.15 0.20 0.25 0.30 Average WD Leaders gpt-4o-mini (hebrew) Qwen2.5-1.5B-Instruct (hebrew) c4ai-command-r7b (hebrew) Phi-3-mini-Instruct (hebrew) Phi-3-medium-Instruct (hebrew) llama-3-8b-Instruct (hebrew) Qwen2.5-7B-Instruct (hebrew) zephyr-7b-beta (hebrew) mistral-7b-v0.3 (hebrew) Gemma3-1b-Instruct (hebrew) //////////////0.144 Qwen2.5-3B-Instruct (hebrew) 0.40 0.05 0.10 0.15 0.20 0.25 0.30 0.35 Average WD Racism \ Antisemitism in the USA Gemma3-1b-Instruct (hebrew) 77,0.286 c4ai-command-r7b (hebrew) gpt-4o-mini (hebrew) zephyr-7b-beta (hebrew) 0.245 mistral-7b-v0.3 (hebrew) Phi-3-medium-Instruct (hebrew) Qwen2.5-1.5B-Instruct (hebrew) Phi-3-mini-Instruct (hebrew) llama-3-8b-Instruct (hebrew) Qwen2.5-3B-Instruct (hebrew) Qwen2.5-7B-Instruct (hebrew) 0.05 0.10 0.20 0.25 0.30 0.35 0.40 0.15 Average WD Terror gpt-4o-mini (hebrew) Gemma3-1b-Instruct (hebrew) 0.110 c4ai-command-r7b (hebrew) 0.104 Qwen2.5-1.5B-Instruct (hebrew) Qwen2.5-7B-Instruct (hebrew) 0.099 0.070 Phi-3-mini-Instruct (hebrew) Phi-3-medium-Instruct (hebrew) 0.065 zephyr-7b-beta (hebrew) llama-3-8b-Instruct (hebrew) 0.062 mistral-7b-v0.3 (hebrew) 0.062 Qwen2.5-3B-Instruct (hebrew) 7// 0.052 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 Average WD Summary 0.353 gpt-4o-mini (hebrew) Gemma3-1b-Instruct (hebrew) Qwen2.5-1.5B-Instruct (hebrew) c4ai-command-r7b (hebrew) zephyr-7b-beta (hebrew) mistral-7b-v0.3 (hebrew) Phi-3-medium-Instruct (hebrew) Phi-3-mini-Instruct (hebrew) Qwen2.5-3B-Instruct (hebrew) llama-3-8b-Instruct (hebrew) Qwen2.5-7B-Instruct (hebrew) 0.20 0.25 0.30 0.35 0.40 0.05 0.10 0.15 Average WD

Model

zephyr-7b-beta

Phi-3-mini-Instruct

c4ai-command-r7b

Gemma3-1b-Instruct

Phi-3-medium-Instruct

Ilama-3-8b-Instruct

mistral-7b-v0.3

gpt-4o-mini

Qwen2.5-1.5B-Instruct

Qwen2.5-3B-Instruct

Qwen2.5-7B-Instruct

Language

english

ZZ hebrew

General Comparison — With Refusal — Hebrew



