Estimated Persuasive Impact in Percentage Points (95% CI)

A

Human

Claude-3-Opus

Qwen1.5-7B

Yi-34B

Falcon-40B

Qwen1.5-4B

GPT-4-Turbo

Llama2-70B

Qwen1.5-14B

Pythia-12B

Pythia-6.9B

Qwen1.5-1.8B

Pythia-1.4B

Pythia-70M

Pythia-410M

10

5

0

0.1   1   10   100   300

Log Scale

B

Qwen1.5-7B

Yi-34B

Qwen1.5-72B

Claude-3-Opus

GPT-4-Turbo

Llama2-70B

Qwen1.5-14B

Pythia-12B

Yi-6B

Pythia-2.8B

Pythia-1.4B

Qwen1.5-0.5B

Pythia-160M

Pythia-70M

0   100   200   300

Linear Scale

Legend	
Falcon	
Frontier	
Llama2	
Pythia	
Qwen1.5	
Yi	

Model Parameters (Billions)