## Signed risk score bias $(SCE_{White} - SCE_{Black})$ Mistral 7B Mixtral 8x7B Llama 38B Mixtral 8x22B Llama 3 70B Yi 34B Mistral 7B (it) Mixtral 8x7B (it) GPT 4o (it) Yi 34B (it) Llama 3 70B (it) GPT 40 mini (it) Base Mixtral 8x22B (it) Instruct Llama 3 8B (it) 0.0 -0.10.1

 $\Delta_{SCF}$