## **Numeric Prompting** Base models Instruction-tuned 1.0 Perf. calibrated Fraction of positives 0.8 Llama 3 70B Mixtral 8x7B 0.6 Mixtral 8x22B Yi 34B 0.4 GPT 40 mini 0.75 1.0 0.0 0.75 0.0 0.25 0.5 1.0 Mean risk score Mean risk score