A Accuracy vs AUROC: yes abst raw logits second prompt, winogrande (r = 70 SOLAR 10.7B Yi 34B xtral 8x7B 65 60 6B Llama2 70B Mistral 7B 55 Q&A Accuracy Llama2 13B 50 Falcon 7B 45 40 35 30 50.0 52.5 55.0 62.5 65.0 67.5 70.0 57.5 60.0 **AUROC**