Q&A Accuracy vs AUROC: yes abst_raw_logits, mmlu (r = 0.84) 0.70 Mixtral 8x7B 0.65 Yi 34B 0.60 SOLAR 10.7B Llama2 70B Mistral 7B 0.55 0.50 0.50 0.45 YI 6B Llama2 13B Falcon 40B Llama2 7B 0.40 0.35 0.30 Falcon 7B 0.650 0.500 0.525 0.550 0.575 0.600 0.625 0.675 0.700

AUROC