Q&A Accuracy vs AUROC: yes abst norm logits, piqa (r = 0.46) Llama 2 lama 2 70 B 4BB SOLAR 10.7B Mixtral 8x7B 0.70 Mistral 7B Falcon 40B 0.65 0.60 0.55 0.50 0.45 Falcon 7B 0.40 0.35 0.30 0.525 0.625 0.650 0.500 0.550 0.575 0.600 0.675 0.700

AUROC