**Performance-Quality Trade-off Matrix (Lower is Better) GPT-2-Large** 0.34 0.34 0.38 0.36 **GPT-2-Medium** 0.41 0.23 0.29 0.50 0.36 **GPT-2-Small** Gemma 7B 0.25 0.45 0.21 0.26 InternLM2.5 7B 0.19 0.28 0.37 0.21 0.37 0.37 0.22 0.32 **LLaMA 2 13B** 0.17 LLaMA 2 13B Chat 0.27 0.32 0.22 0.32 0.76 **LLaMA 2 7B Chat** 0.32 0.28 LLaMA 38B 0.33 0.24 0.04 0.35 model\_display **LLaMA 3.1 8B** 0.27 0.32 0.62 0.39 Mistral 7B 0.14 0.28 0.13 80.0 0.19 0.27 0.19 0.12 **Nous Hermes Mistral 7B DPO** 0.04 0.03 **Nous Hermes Yi 34B** 0.18 0.25 0.14 0.04 OpenChat 3.5 0.15 0.16 OpenChat 3.5 GPTQ 0.31 0.33 0.15 0.33 0.36 0.31 OpenChat 3.5-0106 0.13 0.36 Qwen2 7B 0.34 0.35 0.27 0.31 0.32 StableBeluga 7B 0.32 0.25 0.78 0.12 Yi 6B 0.17 0.18 0.33 Zephyr 7B 0.18 0.10 0.32 0.40 prompt2 prompt1 prompt3 prompt4 prompt

0.7

0.6

0.5

0.4

0.3

0.2

0.1