Harmonic Score per Model and Prompt 0.550 0.517 Nous Hermes Mistral 7B DPO 0.535 0.539 0.541 0.515 0.544 0.498 Zephyr 7B

OpenChat 3.5 GPTQ 0.502 0.502 0.476 0.576 0.575 GPT-2-Large 0.353 0.577 -0.600.438 0.477 0.487 0.602 Mistral 7B 0.575 0.491 OpenChat 3.5-0106 0.475 0.438 OpenChat 3.5 0.501 0.480 0.521 0.477 **-** 0.55 0.508 Nous Hermes Yi 34B 0.431 0.472 0.483 0.340 0.569 0.500 GPT-2-Medium 0.481 0.452 0.581 InternLM2.5 7B 0.299 -0.50Yi 6B 0.367 0.465 0.347 0.630 Model 0.376 0.395 0.481 0.491 Vicuna 13B 0.498 **LLaMA 3.1 8B** 0.437 0.448 0.349 -0.45LLaMA 2 13B 0.424 0.443 0.449 0.377 LLaMA 38B 0.377 0.369 0.432 0.485 0.376 0.503 Qwen2 7B 0.464 0.304 0.664 Mosaic MPT 7B 0.304 0.322 0.336 -0.40LLaMA 2 7B 0.393 0.348 0.384 0.447 0.376 GPT-2-Small 0.390 0.395

- 0.65

StableBeluga 7B 0.411 0.366 0.405 0.313 Gemma 7B 0.352 0.318 0.366 0.448 LLaMA 2 13B Chat 0.340 0.358 0.315 0.465 LLaMA 2 13B Instruct (Exp68) 0.295 0.344 0.323 0.442 prompt1 prompt2 prompt3 prompt4

- 0.35 -0.30Prompt