Model Semantic Behavior Map (T-SNE) gpt2-large gpt-3.5-turbo gpt-neo-1.3B 60 -Llama-2-7b-chat-hf Mistral-7B-instruct-v0.2 Hugging-Face_SmolLM-360M Microsoft_phi-2 40 -Openchat-3.5-1210 StableBeluga-7B Mythomax-I2-13B-Q6 20 tsne_y 0 **-20** · -40 **-60** 20 60 -20 -60 40 -40 tsne_x