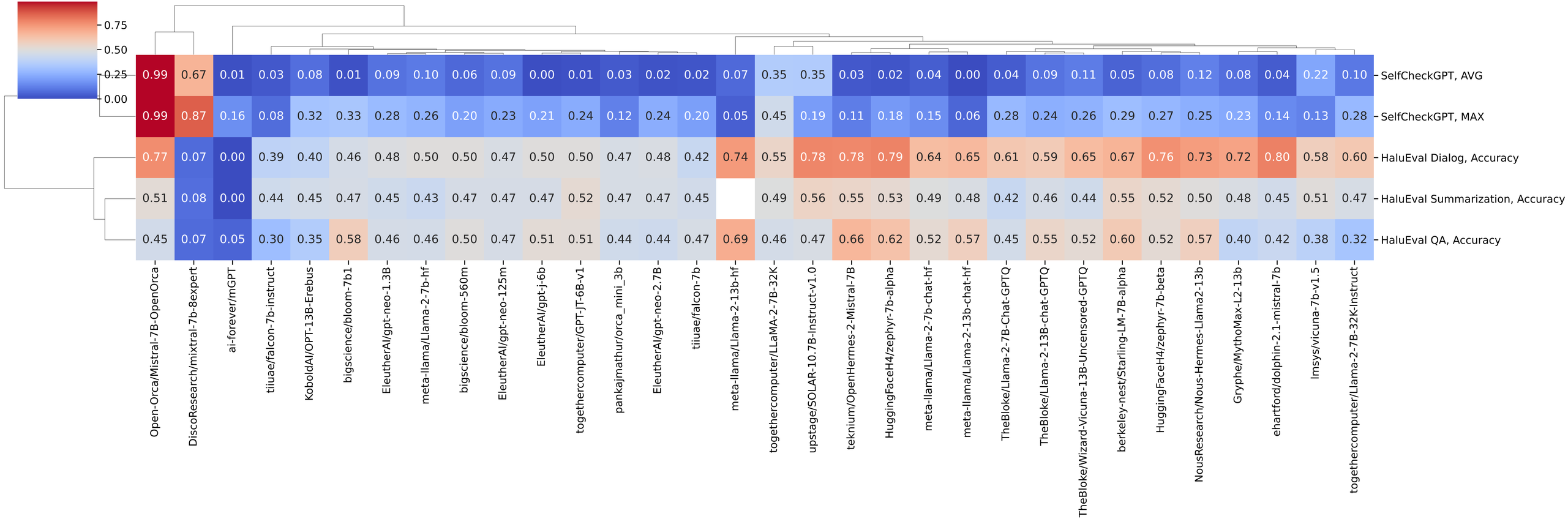
0.75
0.50
0.25
0.00
SelfCheckGPT, AVG
SelfCheckGPT, MAX
HaluEval Dialog, Accuracy
HaluEval Summarization, Accuracy
HaluEval QA, Accuracy
Open-Orca/Mistral-7B-OpenOrca
DiscoResearch/mixtral-7b-8expert
ai-forever/mGPT
tiiuae/falcon-7b-instruct
KoboldAI/OPT-13B-Erebus
bigscience/bloom-7b1
EleutherAI/gpt-neo-1.3B
meta-llama/Llama-2-7b-hf
bigscience/bloom-560m
EleutherAI/gpt-neo-125m
EleutherAI/gpt-j-6b
togethercomputer/GPT-JT-6B-v1
pankajmathur/orca_mini_3b
EleutherAI/gpt-neo-2.7B
tiiuae/falcon-7b
meta-llama/Llama-2-13b-hf
togethercomputer/LLaMA-2-7B-32K
upstage/SOLAR-10.7B-Instruct-v1.0
teknium/OpenHermes-2-Mistral-7B
HuggingFaceH4/zephyr-7b-alpha
meta-llama/Llama-2-7b-chat-hf
meta-llama/Llama-2-13b-chat-hf
TheBloke/Llama-2-7B-Chat-GPTQ
TheBloke/Llama-2-13B-chat-GPTQ
TheBloke/Wizard-Vicuna-13B-Uncensored-GPTQ
berkeley-nest/Starling-LM-7B-alpha
HuggingFaceH4/zephyr-7b-beta
NousResearch/Nous-Hermes-Llama2-13b
Gryphe/MythoMax-L2-13b
ehartford/dolphin-2.1-mistral-7b
lmsys/vicuna-7b-v1.5
togethercomputer/Llama-2-7B-32K-Instruct