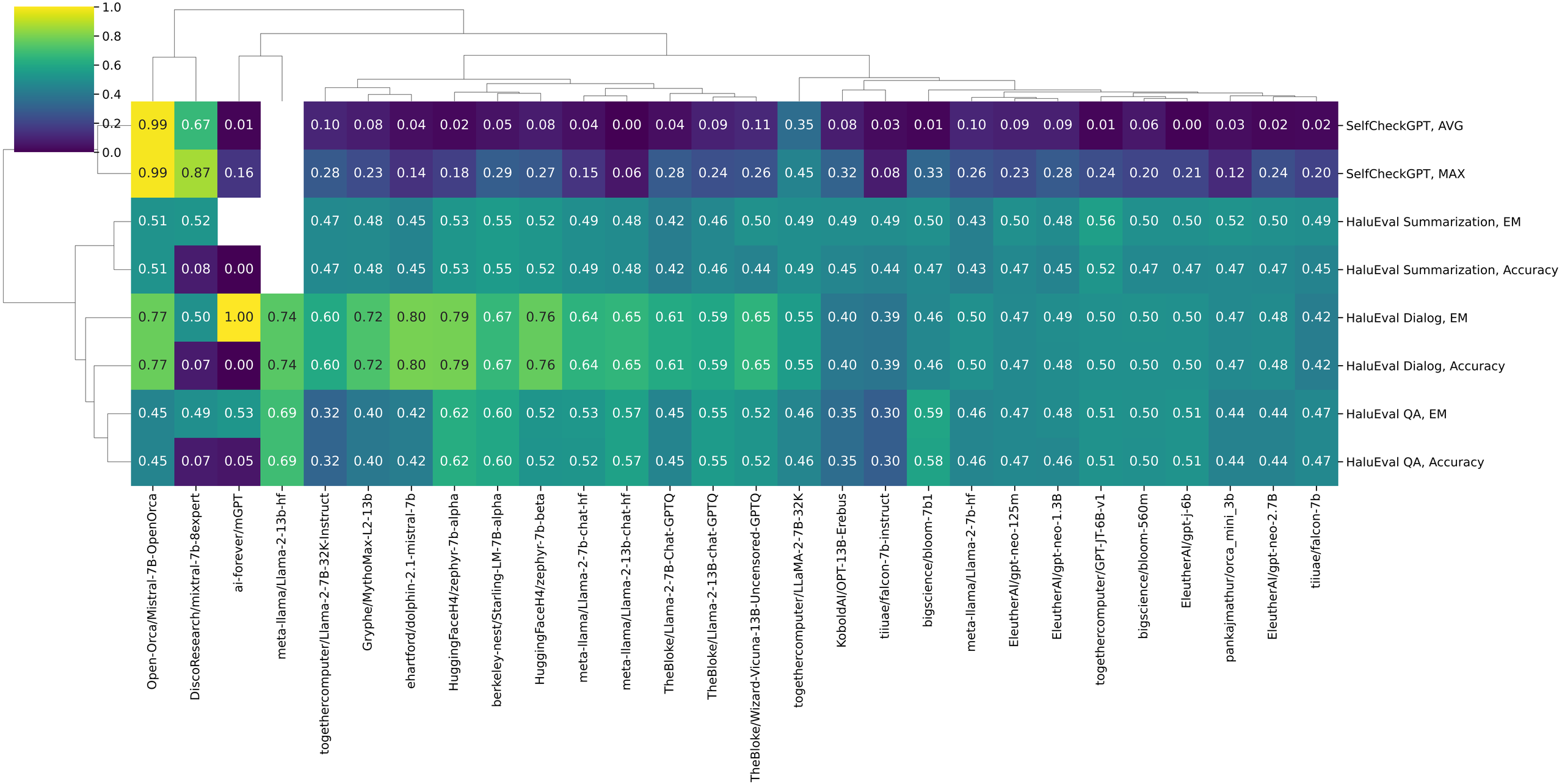
1.0
0.8
0.6
0.4
0.2
0.0
SelfCheckGPT, AVG
SelfCheckGPT, MAX
HaluEval Summarization, EM
HaluEval Summarization, Accuracy
HaluEval Dialog, EM
HaluEval Dialog, Accuracy
HaluEval QA, EM
HaluEval QA, Accuracy
Open-Orca/Mistral-7B-OpenOrca
DiscoResearch/mixtral-7b-8expert
ai-forever/mGPT
meta-llama/Llama-2-13b-hf
togethercomputer/Llama-2-7B-32K-Instruct
Gryphe/MythoMax-L2-13b
ehartford/dolphin-2.1-mistral-7b
HuggingFaceH4/zephyr-7b-alpha
berkeley-nest/Starling-LM-7B-alpha
HuggingFaceH4/zephyr-7b-beta
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
togethercomputer/LLaMA-2-7B-32K
KoboldAI/OPT-13B-Erebus
tiiuae/falcon-7b-instruct
bigscience/bloom-7b1
meta-llama/Llama-2-7b-hf
EleutherAI/gpt-neo-125m
EleutherAI/gpt-neo-1.3B
togethercomputer/GPT-JT-6B-v1
bigscience/bloom-560m
EleutherAI/gpt-j-6b
pankajmathur/orca_mini_3b
EleutherAI/gpt-neo-2.7B
tiiuae/falcon-7b
0.99 0.67 0.01 0.10 0.08 0.04 0.02 0.05 0.08 0.04 0.00 0.04 0.09 0.11 0.35 0.08 0.03 0.01 0.10 0.09 0.09 0.01 0.06 0.00 0.03 0.02 0.02
0.99 0.87 0.16 0.28 0.23 0.14 0.18 0.29 0.27 0.15 0.06 0.28 0.24 0.26 0.45 0.32 0.08 0.33 0.26 0.23 0.28 0.24 0.20 0.21 0.12 0.24 0.20
0.51 0.52 0.47 0.48 0.45 0.53 0.55 0.52 0.49 0.48 0.42 0.46 0.50 0.49 0.49 0.49 0.50 0.43 0.50 0.48 0.56 0.50 0.50 0.52 0.50 0.49
0.51 0.08 0.00 0.47 0.48 0.45 0.53 0.55 0.52 0.49 0.48 0.42 0.46 0.44 0.49 0.45 0.44 0.47 0.43 0.47 0.45 0.52 0.47 0.47 0.47 0.47 0.45
0.77 0.50 1.00 0.74 0.60 0.72 0.80 0.79 0.67 0.76 0.64 0.65 0.61 0.59 0.65 0.55 0.40 0.39 0.46 0.50 0.47 0.49 0.50 0.50 0.50 0.47 0.48 0.42
0.77 0.07 0.00 0.74 0.60 0.72 0.80 0.79 0.67 0.76 0.64 0.65 0.61 0.59 0.65 0.55 0.40 0.39 0.46 0.50 0.47 0.48 0.50 0.50 0.50 0.47 0.48 0.42
0.45 0.49 0.53 0.69 0.32 0.40 0.42 0.62 0.60 0.52 0.53 0.57 0.45 0.55 0.52 0.46 0.35 0.30 0.59 0.46 0.47 0.48 0.51 0.50 0.51 0.44 0.44 0.47
0.45 0.07 0.05 0.69 0.32 0.40 0.42 0.62 0.60 0.52 0.52 0.57 0.45 0.55 0.52 0.46 0.35 0.30 0.58 0.46 0.47 0.46 0.51 0.50 0.51 0.44 0.44 0.47