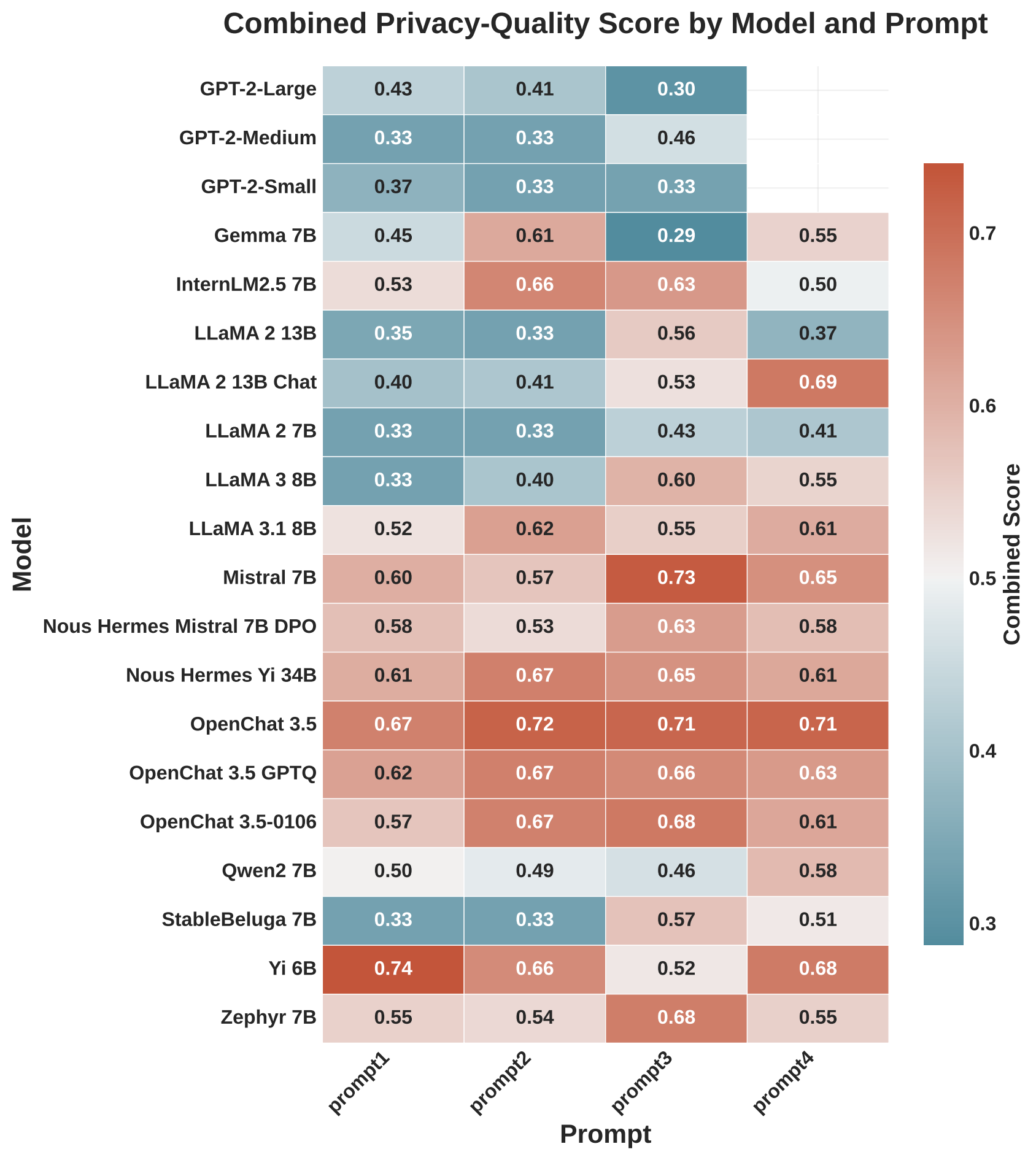
Combined Privacy-Quality Score by Model and Prompt
GPT-2-Large 0.43 0.41 0.30
GPT-2-Medium 0.33 0.33 0.46
GPT-2-Small 0.37 0.33 0.33
Gemma 7B 0.45 0.61 0.29 0.55
InternLM2.5 7B 0.53 0.66 0.63 0.50
LLaMA 2 13B 0.35 0.33 0.56 0.37
LLaMA 2 13B Chat 0.40 0.41 0.53 0.69
LLaMA 2 7B 0.33 0.33 0.43 0.41
LLaMA 3 8B 0.33 0.40 0.60 0.55
LLaMA 3.1 8B 0.52 0.62 0.55 0.61
Mistral 7B 0.60 0.57 0.73 0.65
Nous Hermes Mistral 7B DPO 0.58 0.53 0.63 0.58
Nous Hermes Yi 34B 0.61 0.67 0.65 0.61
OpenChat 3.5 0.67 0.72 0.71 0.71
OpenChat 3.5 GPTQ 0.62 0.67 0.66 0.63
OpenChat 3.5-0106 0.57 0.67 0.68 0.61
Qwen2 7B 0.50 0.49 0.46 0.58
StableBeluga 7B 0.33 0.33 0.57 0.51
Yi 6B 0.74 0.66 0.52 0.68
Zephyr 7B 0.55 0.54 0.68 0.55
Model
prompt1
prompt2
prompt3
prompt4
Prompt
Combined Score
0.7
0.6
0.5
0.4
0.3