MT-Bench vs YokaiEval

Model Category
- Chinese-centric (green triangle)
- Japanese-centric (red circle)
- English-centric (blue square)

gpt4o-mini
gpt4o
meta-llama-3.3-70b-instruct
llama-3.1-70b-japanese-instruct-2407
llama-3.1-swallow-70b-instruct-v0.1
llama-3-elyza-jp-8b
calm3-chat
qwen2.5-32b-instruct
qwen2.5-7b-instruct
llama-3.1-swallow-8b-instruct-v0.3
llama-3.1-swallow-8b-instruct-v0.1
qwen2-7b-instruction
gemma-2-27b-it
qwen2.5-14b-instruct
qwen2.5-3b-instruct
llama-3.1-swallow-8b-instruct-v0.2
meta-llama-3.1-8b-instruct
meta-llama-3-8b-instruct
tanuki-8x8b-dpo-v1.0
swallow-13b-instruct-v0.1
mistral-nemo-instruct-2407
gemma-2-9b-it
llm-jp-3-13b-insturct
llama-3.2-3b-instruct
gemma-2-2b-it
llm-jp-3-1.8b-instruct
llm-jp-3-3.7b-instruct
mistral-7b-instruct-v0.2
tanuki-8b-dpo-v1.0

MT-Bench
YokaiEval