KL Differences Relative to OPT 2.7B - * Larger Datasets *

Legend:
- Gemma2 2B (8888GB)
- Qwen2.5 0.5B (80000GB)
- Qwen2.5 3B (80000GB)
- OPT 2.7B (800GB)