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

Legend:
- GPT2 (40GB)
- GPT2 Large (40GB)
- OPT 6.7B (800GB)
- OPT 2.7B (800GB)