KL Differences Relative to GPT2 Large - * Larger Datasets *