

KL Differences Relative to GPT2 - * Later Models *

