

PPL Variation over First 1500 Iterations on 345M GPT with GPT-2-Output-Dataset

