

Average Sequence Perplexity heatmaps with $\log(1 - p)$ logit control