

Temperature Calculation: Step-by-Step

Given: $\text{Logits} = [2.0, 1.0, 0.5, 0.2]$
 $\text{Tokens} = [\text{"cat"}, \text{"dog"}, \text{"bird"}, \text{"fish"}]$

Step 1: Scale by $T=0.5$ (MORE UNBALANCED)
 $\text{Scaled} = [4.0, 2.0, 1.0, 0.5]$
 $\text{Softmax} \rightarrow [0.63, 0.19, 0.12, 0.05]$
 $\rightarrow 73\% \text{ on "cat"} \text{ (VERY FOCUSED)}$

Step 3: Scale by $T=2.0$ (FLATTER)
 $\text{Scaled} = [1.0, 0.5, 0.25, 0.125]$
 $\text{Softmax} \rightarrow [0.38, 0.17, 0.15, 0.09]$
 $\rightarrow \text{CH FLATTER}$

Lower $T \rightarrow$ More confident (peaky)
Higher $T \rightarrow$ More random (flat)

