

Temperature Calculation: Step-by-Step

Given: Logits = [2.0, 1.0, 0.5, 0.2]
Tokens = ["cat", "dog", "bird", "fish"]

Step 1: Scale Scaled by 0.5 (MORE FOCUSED)
Scaled = $\frac{[2.0, 1.0, 0.5, 0.2]}{\sqrt{0.5}} = [4.0, 2.0, 1.0, 0.4]$
Softmax = $\frac{e^{[4.0, 2.0, 1.0, 0.4]}}{\sum e^{[4.0, 2.0, 1.0, 0.4]}} = [0.53, 0.19, 0.12, 0.10]$
→ 73% on "cat" (**VERY FOCUSED**)

Step 3: Scale by T=2.0 (FLATTER)

$$\text{Softm } p_i = \frac{\exp(\text{logit}_i/T)}{\sum \exp(\text{logit}_i/T)}$$

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