

Temperature Calculation: Step-by-Step

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

Step 1: Scale by $T=0.5$ (MORE UNBALANCED)
Scaled Logits = [4.0, 2.0, 1.0, 0.5]
Softmax → [0.63, 0.19, 0.12, 0.10]
→ 73% on "cat" (VERY FOCUSED)

Step 3: Scale by $T=2.0$ (FLATTER)
Scaled Logits = [1.0, 0.5, 0.25, 0.125]
Softmax → [0.38, 0.17, 0.15, 0.15]
→ 38% on "cat" (CH FLATTER)

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

