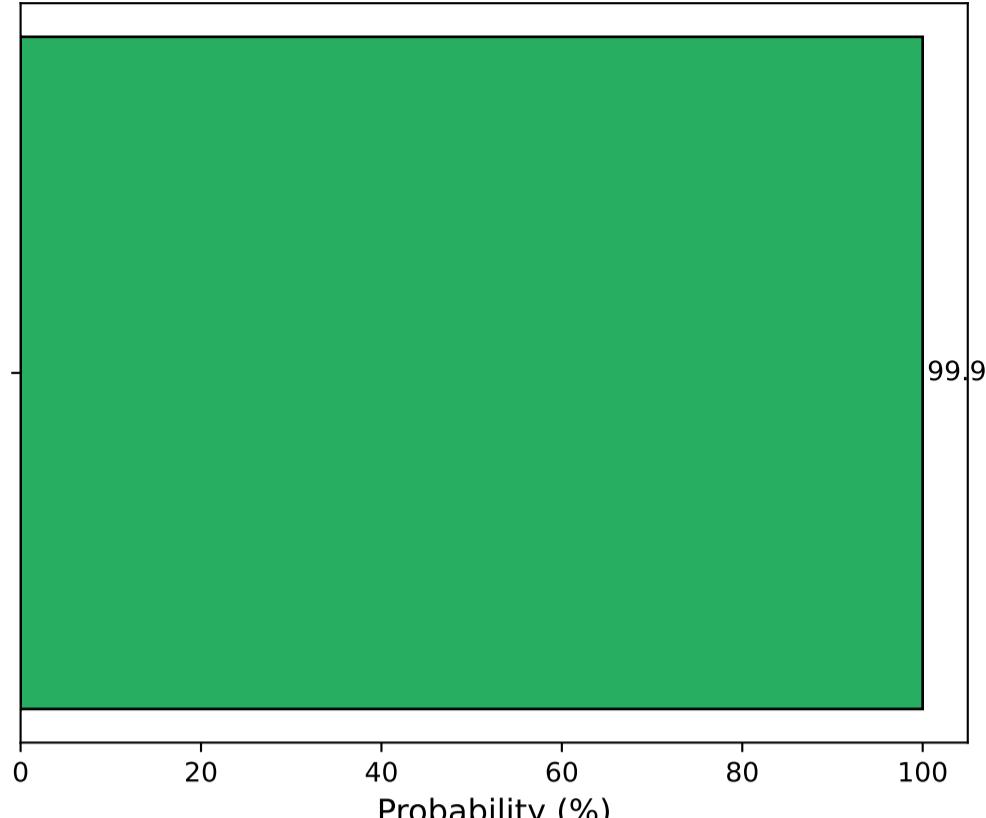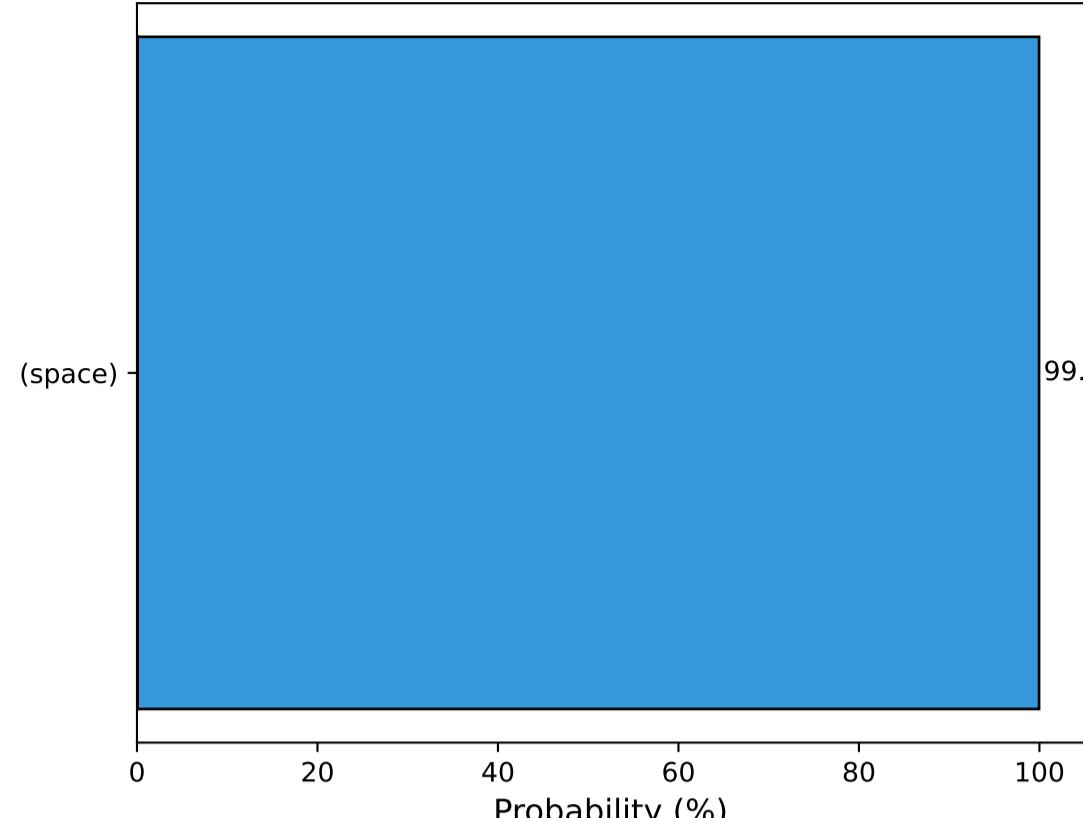**Figure 3: Token Probability Distribution Comparison**

TinyLlama (Original)

TinyLlama (Ours - Modified)

The /0.8 modification sharpens probability distribution, making the model more deterministic.