Bigram probability of "They play in big garden"

Use MLE bigrams with a start token `<s>`:

$$P(\text{They play in big garden}) = P(\text{They} \mid \langle s \rangle)\, P(\text{play} \mid \text{They})\, P(\text{in} \mid \text{play})\, P(\text{big} \mid \text{in})\, P(\text{garden} \mid \text{big})$$

From the training set (tokenized, lowercased for counting):

1. $P(\text{They} \mid \langle s \rangle) = \dfrac{\text{Count}(\langle s \rangle, \text{They})}{\text{Count}(\langle s \rangle)} = \dfrac{1}{3}$
2. $P(\text{play} \mid \text{They}) = \dfrac{\text{Count}(\text{They, play})}{\text{Count}(\text{They})} = \dfrac{1}{1} = 1$
3. $P(\text{in} \mid \text{play}) = \dfrac{\text{Count}(\text{play, in})}{\text{Count}(\text{play})} = \dfrac{1}{2}$
   - ("play in" once; "play" occurs twice: "play in...", "play inside...")
4. $P(\text{big} \mid \text{in}) = \dfrac{\text{Count}(\text{in, big})}{\text{Count}(\text{in})} = \dfrac{0}{1} = 0$
   - ("in big" never occurs; only "in the")
5. $P(\text{garden} \mid \text{big}) = \dfrac{\text{Count}(\text{big, garden})}{\text{Count}(\text{big})} = \dfrac{1}{1} = 1$

Now multiply:

$$\left(\frac{1}{3}\right) \times (1) \times \left(\frac{1}{2}\right) \times (0) \times (1) = \boxed{0}$$

Result: Probability is **0** (unsmoothed bigram assigns zero due to unseen bigram "in big").

(With smoothing, this would become small but non-zero.)

AI ML FREE ROADMAP

Maths
Python
LIB - Numpy, Pandas, Matplotlib
ML Fundamentals
Frameworks - Scikit Learn

Deeplearning
Frameworks - Tensorflow
GenAI
Vector DB
RAG
Lang Chain
AI Agents