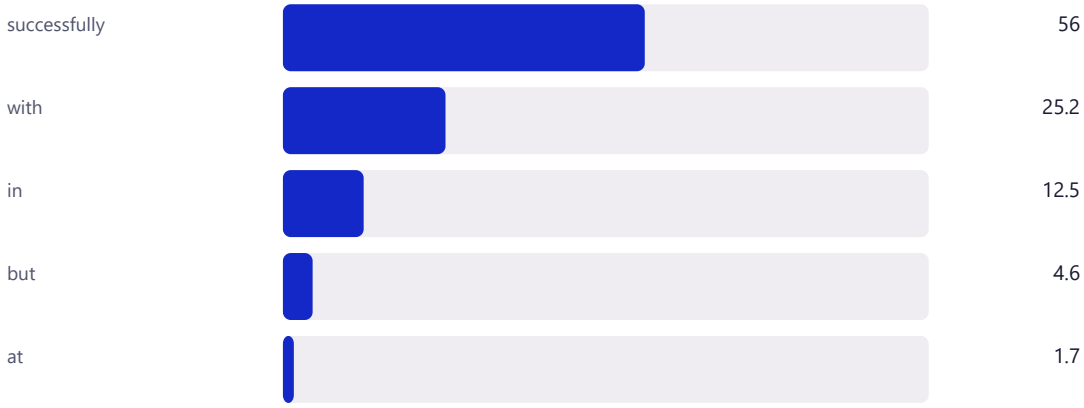
CHAPTER 09 / SAMPLING

One token. One decision.

Reshape scores, filter the tail, then choose.

Reshape scores, filter the tail, then choose.



Handpicked example values · inspect exact values below

Five handpicked logits and seeded samples. Probabilities are not truth scores or a hosted-service reproducibility promise.

Scores

Softmax turns logits divided by temperature into a distribution.

Filters

Top-k limits count; top-p keeps the smallest ranked set reaching the chosen mass.

Choice

Renormalize the retained set and choose a token. The choice becomes next-step context.

↗ **Softmax turns logits divided by temperature into a distribution.**

PUT IT TO WORK

Try top-k 2, then top-p 0.9. The model weights stay unchanged while the allowed choices change.

CLOSE THE PAGE. EXPLAIN IT.

Which settings change the selection rule rather than the model weights?