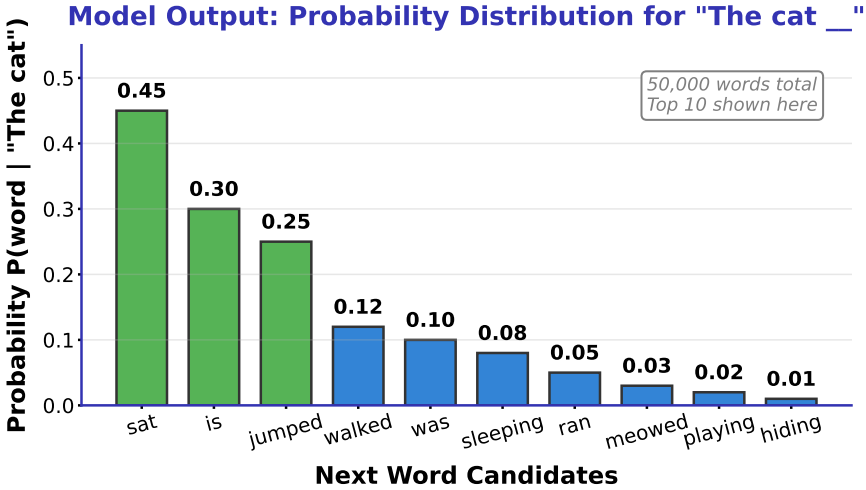
PREVIEW: Choose Your Approach

4 Setup Charts + 4 Problem Framings

November 9, 2025

4 ways to show “The cat __” probability problem

Next 4 slides show all options



Pros: Visual, shows distribution clearly
Cons: Takes more space

Setup Option B: List Format

Model Output for: "The cat __"

P(sat) = 0.45
P(is) = 0.30
P(jumped) = 0.25
P(walked) = 0.12
P(was) = 0.10

Top 3 candidates (85% of probability mass)

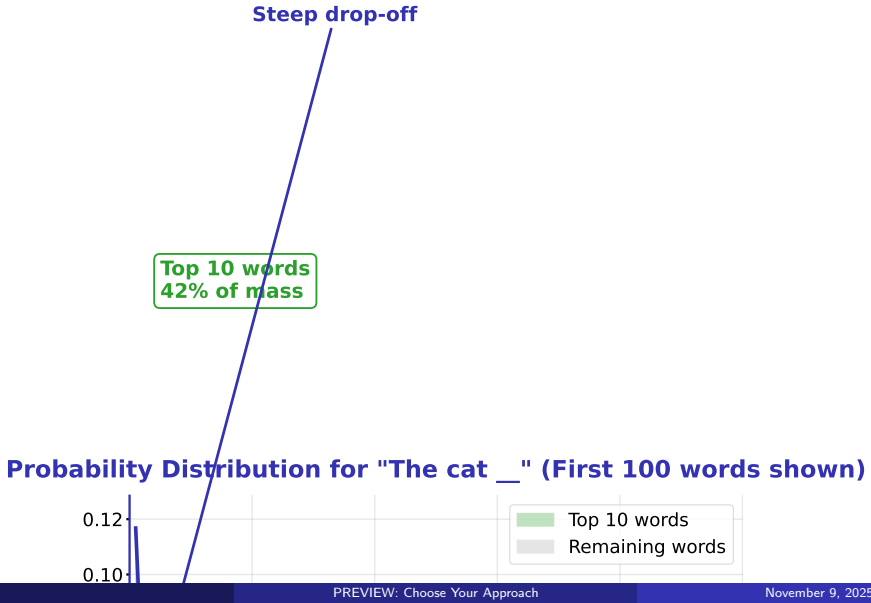
P(sleeping) = 0.08
P(ran) = 0.05
P(meowed) = 0.03
...

Total vocabulary: 50,000 words

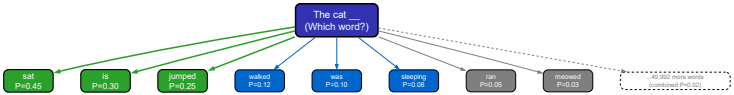
Challenge: Which word do we pick?

Pros: Clean, simple, text-focused
Cons: Less visual

Setup Option C: Distribution Curve



Setup Option D: Node Diagram



Pros: Decision tree style, shows choices
Cons: More complex layout

Frame as: "What goes wrong and how to fix it"

6 failure modes $\rightarrow$ 6 solutions

Next 6 slides show this framing

Problem:

Problem 1: Repetitive Loops

Framing 1: Failure Modes

Solution: **BEAM SEARCH**
Explores multiple paths
Finds better sequences
Width = 3-5 typical

Avoids local optima

"The city is a major city in a city..."
Greedy gets stuck in loops

Problem:

Problem 2: No Diversity

Framing 1: Failure Modes

Always same output
Deterministic = boring

Solution: **TEMPERATURE**
Adds randomness
T $<$ 1: focused
T $>$ 1: creative

Continuous control over creativity

Problem:

Problem 3: Tail Sampling

Framing 1: Failure Modes

Solution: TOP-K
Filter bottom 99%
Sample from top-k only
k = 40-50 typical

Prevents nonsense sampling

Random samples unlikely words
"purple flying mathematics"

Problem:

Problem 4: Fixed Cutoff

Framing 1: Failure Modes

Solution: **NUCLEUS (TOP-P)**
Dynamic cutoff
Adapts to distribution
$p = 0.9$ typical

Smart filtering

Top-k inflexible
Same k for peaked and flat

Problem:

Problem 5: Long Degeneration

Framing 1: Failure Modes

Long texts repeat more
Models copy recent context

Solution: CONTRASTIVE
Penalize similarity
$\alpha = 0.6$ typical
Explicit degeneration prevention

Best for long generation

Problem:

Problem 6: Quality-Diversity Tradeoff

Framing 1: Failure Modes

Solution: **CHOOSE BY TASK**
Translation → Beam
Creative → Nucleus/Contrastive
No universal best

Task determines method

Cannot maximize both
Fundamental tradeoff

Frame as: "What do we need and which method provides it"

6 needs $\rightarrow$ 6 methods

Next 6 slides show this framing

Framing 2: 6 Needs (Summary)

1. Need exact/correct → GREEDY
2. Need better sequences → BEAM
3. Need creativity → TEMPERATURE
4. Need filter unlikely → TOP-K
5. Need adapt to distribution → NUCLEUS
6. Need prevent repetition → CONTRASTIVE

Pedagogical approach: Positive framing (what we want to achieve)
vs Framing 1 (what goes wrong)

Frame as: "Building up solutions progressively"

Start simple $\rightarrow$ Add sophistication $\rightarrow$ Solve harder problems

Shows natural evolution of methods

Framing 3: Progressive Levels (Summary)

1. Level 1: Greedy too deterministic → Add Temperature
2. Level 2: Temperature too random → Add Top-k filtering
3. Level 3: Top-k inflexible → Dynamic Nucleus
4. Level 4: Still miss good paths → Beam exploration
5. Level 5: Still repetitive (long) → Contrastive penalty
6. Level 6: Need hybrid → Combine methods

Pedagogical approach: Scaffolding (each method fixes previous limitation)

Frame as: "Here's actual bad output, here's the fix"

6 concrete failures $\rightarrow$ 6 targeted solutions

Most concrete/motivating approach

Framing 4: Bad Outputs (Summary)

1. Output: "city is a city in a city" → FIX: Contrastive
2. Output: "purple flying mathematics" → FIX: Top-k/Nucleus
3. Output: Always same (boring) → FIX: Temperature
4. Output: Missed better sequence → FIX: Beam
5. Output: Wrong for distribution type → FIX: Nucleus
6. Output: Too slow for production → FIX: Greedy

Pedagogical approach: Concrete examples first (most engaging)

RECOMMENDATIONS

My Recommendation:

Best Framing: Framing 4 (Real Bad Outputs)
Why:

- Most concrete and motivating
- Students see actual failures
- Each method clearly solves specific problem
- Memorable (students remember the bad examples)

Best Setup Chart: Option A (Bar Chart)
Why:

- Visual and quantitative
- Shows distribution clearly
- Easy to understand at a glance

Structure: 6 compact problem-solution slides + hybrid details (9 slides)
Total Main: 20 slides (very focused and effective)