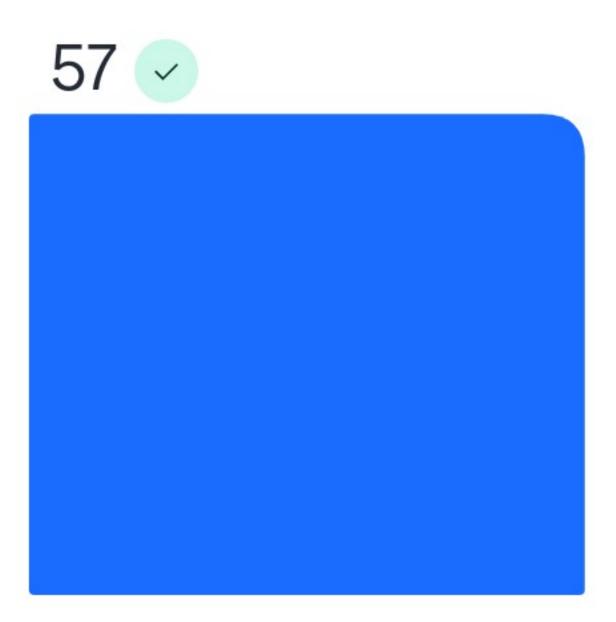
Which of the following are limitations of statistical language models that neural language models help solve?



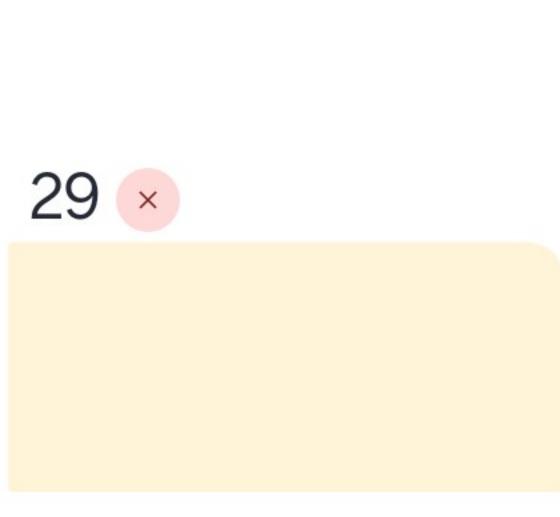
Conditioning on long contexts is prohibtively expensive.



They cannot handle non-English characters.



Sequences that were unseen during training are assigned zero probability.

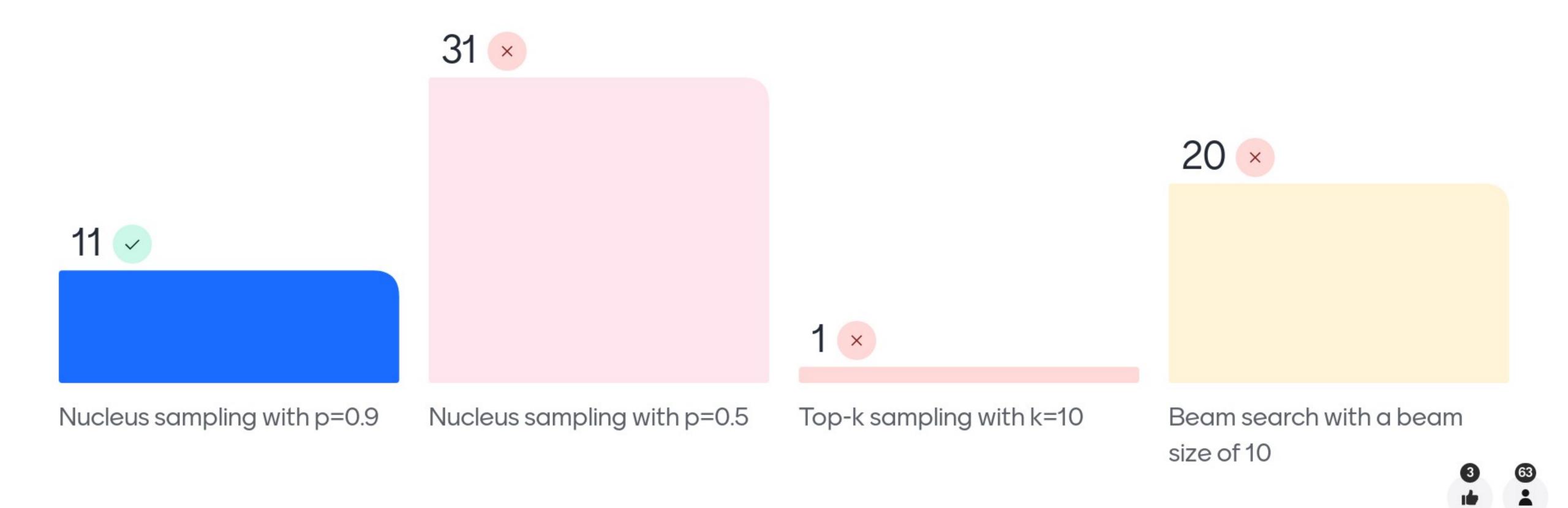


They are extremely slow to compute sequence likelihoods.





Which of the following decoding strategies will result in the lowest likelihood generated text?



How do most subword tokenizers handle words that are not in their vocabulary?



Break the word down into substrings which are in the vocabulary

2 ×

Break the word down into individual characters.



Replace the word with a special token, indicating it's unknown



Skip tokenizing this word



