CS685 Quiz 1: *Neural language models*
Released 2/20, due 2/27 on Gradescope (please upload a PDF!)
Please answer both questions in 2-4 sentences each. Make sure to also fill out the AI disclosure!

1. Explain what the “bottleneck” of a recurrent neural network is and how attention provides a way to get around this bottleneck.

   The “bottleneck” of a recurrent neural network refers to how in order to calculate the hidden representation at a given time step, you need the hidden representation of the previous time step, which means you have to finish a calculation before moving on to the next. The attention provides a way to get around this bottleneck because it parallelizes the calculation of the hidden representations.

2. You are given two language models trained on Wikipedia. One is an unsmoothed 5-gram model (i.e., prefixes are four tokens long), while the other is a fixed-window neural language model with an identical prefix size. Which model’s estimate of the conditional probability distribution P(*w* | “*chalkboards flap their wings*”) is likely to be more reasonable and why?

Unsmoothed 5-gram model is more likely. Neural language models are trained on patterns on human text datasets and the phrase is very unlikely to occur there.

AI Disclosure

AI1: Did you use any AI assistance to complete this homework? If so, please also specify what AI you used.
Your answer here

(only complete the below questions if you answered yes above)

AI2: If you used a large language model to assist you, please paste *all* of the prompts that you used below. Add a separate bullet for each prompt, and specify which problem is associated with which prompt.

- *Your response here*

AI3: (*Free response)* For each problem for which you used assistance, describe your overall experience with the AI. How helpful was it? Did it just directly give you a good answer, or did you have to edit it? Was its output ever obviously wrong or irrelevant? Did you use it to get the answer or check your own answer?

- *Your response here*