# Exploring Language Models

One person on each team should log in to the OpenAI Playground:
https://platform.openai.com/playground

Objectives:

- Describe the implications of how language models generate text sequentially.
- Describe what a conditional distribution is, in the context of language modeling.
- Compute the log-probability that a language model assigns to of a sequence of words.

## Part 1: Left-to-Right Generation

1. Type this into the Playground: "The number 17433 is composite because it can be written as the product of". Don't type a space afterwards. Leave all parameters at their defaults. Click Submit to generate. (It should give several numbers; if not, try again.) Check its output using a calculator. Is it correct?

2. Repeat the previous step a few more times. (The "Regenerate" button makes this easy.) Keep track of what factorizations it generates and whether they are correct:

3. Now change the prompt to "The number 17433 is prime because" and generate. What do you notice?

## Part 2: Token Probabilities

4. Use the prompt "Tell me a joke." Set the Temperature slider to 0. What joke is generated?

5. Compare your response to the previous question with that of a neighboring team. What do you notice?

6. Now set the Temperature slider to 1 and Regenerate. What joke is generated?

7. Repeat the previous step a few times. Summarize what you observe.

8. Under “Show probabilities”, select “Full spectrum” (you’ll need to scroll down). Generate with a temperature of 0 again. Select the initial “Q”; you should see a table of words with corresponding probabilities. What options was the model considering for how to start the joke?

9. Click each word in the generated text. (Make sure it was generated with Temperature set to 0.) Notice the words highlighted in red; those are the words that were chosen from the conditional distribution. How do you think the model chooses from among the options it’s considering when Temperature is 0?

10. Now set Temperature to 1 and Regenerate. How do you think the model chooses from among the options it’s considering when Temperature is 1? Regenerate a few times to check your reasoning.

11. Observe the highlighting behind each word. Describe what it means when a token is red.

12. Suppose the logits for two words are 0.1 and 0.2. Softmax these logits to obtain probabilities. Now divide the logits by .001 and again compute the softmax. The number you divide the logits by is the *temperature*.

## Part 3: Phrase Probabilities

13. Select the first few words of the generated joke. You should see “Total: xx.xxx logprob on yy tokens”. Write down the logprob number.

14. Click the first token and observe the corresponding “Total:” statement for that token. Write down the logprobs reported individually for each token, for the first few tokens.

15. Sum the logprobs of each token. Check that the sum of the individual token logprobs matches the total logprob reported for the phrase.

16. Compute the logprob for one token by computing the natural logarithm of the probability of the chosen word.

17. Briefly describe the relationship of these logprobs to *cross-entropy*.