activity: act out an LLM

Learning targets

Students will be able to:

- Define **token** and explain the importance of tokens in LLMs.
- Explain how LLMs produce human-sounding text.
- Explain concerns about **hallucinations** and biased training data in LLMs.

Students will...

- Define the term **token** and compare with counterexamples.
- Act out the process of an LLM generating text based on a prompt.
- Discuss concerns about hallucinations in LLMs.

Key vocab

- Token: A piece of a sentence that contains meaning. It can be a word, part of a word, or a punctuation mark.
- Hallucination: An incorrect piece of text generated by an LLM.

Prep

- Decide on a voting mechanism for the activity. Options include:
  - ↳ Have students write on personal whiteboards
  - ↳ Use an online polls (e.g., Google Forms)
  - ↳ Have students vote on the classroom whiteboard using sticky notes

---

intro | whole class | 3-5 min

Preview the day's learning targets with students.

Define the word **token** and explain the examples and counterexamples. Tokens should be long enough that they are meaningful but not so long that they could be broken up further.

vocab spotlight

Token: A piece of a sentence that contains meaning. It can be a word, part of a word, or a punctuation mark.

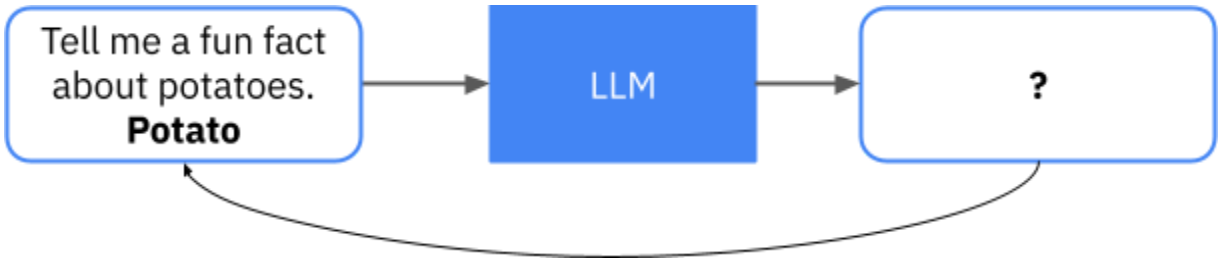
act out an LLM | whole class | 10-12 min

Introduce the goal of the activity: to act out the process of an LLM generating a sentence.

Explain the diagram below.

We give a prompt to an LLM, which breaks the prompt into tokens. The LLM then considers all possible tokens from its token bank and selects the one most likely to be the starting word of the response. In this case, that token is "Potato."

Explain further that the generated token is added to the end of the prompt and fed back into the LLM:



Prompt students to generate the following tokens by voting on what token from the bank should be generated next.

Token bank

A	America	and	are	part	potato	root	staple
as	Asia	crop	first	starch	the	to	were
food	for	grown	it	-s	-es	-ing	-y

Continue repeating the process of generating a token, then adding it to the prompt. Repeat until a full sentence is generated.

facilitator note

Generally, you'll want to select the token with the most votes from students. However, LLMs often use some randomness when determining the next token to select. Consider picking the second- or third- most popular choice a few times during the activity.

Define **hallucination** and ask students:

- Is there a way to know if the sentence we produced was correct?

vocab spotlight

Hallucination: An incorrect piece of text generated by an LLM.

references

This activity is adapted from part of Melanie Mitchell's talk, *The Future of Artificial Intelligence* (21:57-23:16). https://www.youtube.com/watch?v=HAiXT1mGTXc&t=313s.

potential activity extensions

These activities surround the “Act Out An LLM” activity and contain an additional learning target:

- Explain the role of **attention mechanisms** in LLMs.

warm-up | whole class | 2-4 min

Prompt the class to shout out and fill in the blanks for each phrase:

- Once upon a _____
- Happy birthday to _____
- My favorite color is _____
- Over the weekend, I _____
- The _____

Ask students: Why are some phrases easier to predict than others?

attention mechanisms | pairs| 8-10 min

Explain the role of an **attention mechanism** in an LLM.

vocab spotlight

Attention mechanism: The part of an LLM that finds connections between tokens in order to predict what should be generated.

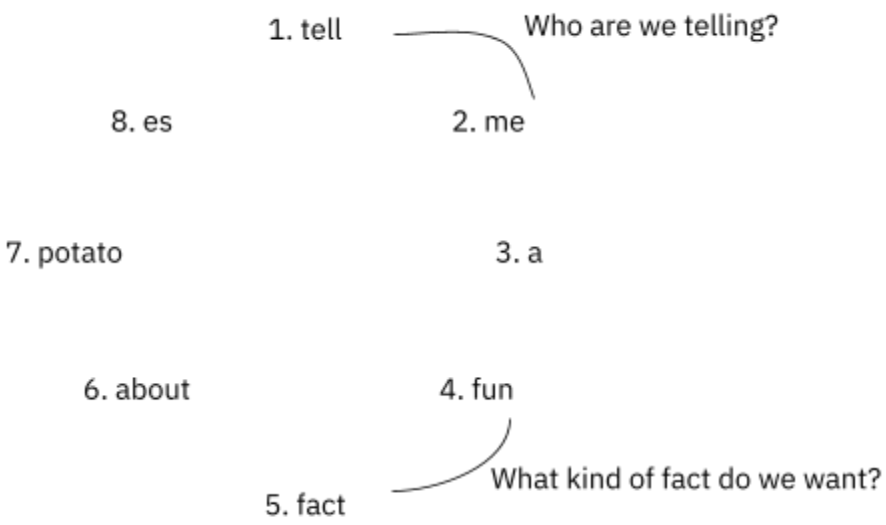
Given the example prompt “Tell me a fun fact about potatoes,” direct pairs of students to find two tokens in the sentence that are meaningfully connected.

Prompt students to share their findings with the whole class, optionally drawing connections on the board to represent students’ suggestions.

Connections may include:

Token	Token	Reason
tell	me	Who are we telling?
me	a	Neighbors in the sentence

a	fact	Filling in the blank: a (what)
about	potato	Defining what the fact is about
potato	es	Make up a word



Example connections made by students.

The tokens are numbered to represent their positional connection: "tell" and "me" are meaningfully connected in the sentence because they're neighbors.

responsible design: token dictionaries | whole class | 3-5 min

Explain: LLMs read a lot of text; over 500 billion words. Their dictionaries contain over 100,000 tokens!

When an LLM is trained, the designers give it chunks of text from the internet to "read" and process.

Ask students:

- For our potato example, what are possible websites our example dictionary was trained on?
- How would our LLM be different if it trained only on mystery novels? What about websites containing a lot of toxic language?

exit ticket | individual | 3-5 min

Prompt students to respond to the exit ticket:

1. What is a **token** in the context of an LLM?
2. What does an **attention mechanism** do in an LLM?
3. What does it mean for an LLM to **hallucinate**?
4. LLMs hallucinate often and can be biased in their training data. How does knowing that change the way we should use LLMs?