

I chose this word because::

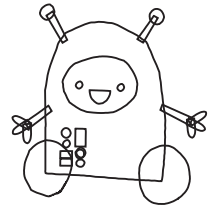
My word:

Today you'll add one word to our shared word space. Which word will you choose?

Think about your morning: what did you touch today? · Do you know a word in another language? · Mix two words and invent a new one!

## Choose your word

### Task 1:



## Adventure in Word Space

Have you ever chatted with a computer? Asked a phone to tell a joke or make up a story?

Machines only learned to speak and write fluently just now. But how did a machine learn? Well — by reading. Almost everything humans have ever written: stories, letters, recipes, internet posts... From all that reading, people built a language model: an exciting new material, a bit like clay or crayons, but made of words.

**Training data.** The enormous pile of text the machine learned from — more than in any library in the world.



**Token.** A piece of text the machine reads and writes with. The machine's alphabet has about 50,000 tokens.

Chop your own word into pieces below:

A language model doesn't read letters or whole words — it chops words into pieces. Pieces that appear often in texts earned a place in the machine's alphabet. That's why a familiar word is one piece, but a rare word gets chopped:

## The machine reads in pieces

### Task 2:

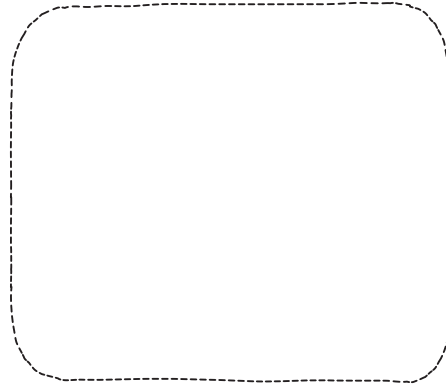
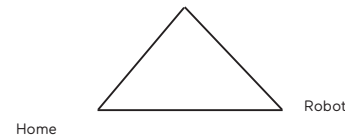
## Task 7: Add your word to the language model

Now you know all kinds of things about how language models work. It's time to add your own word. Write or stamp your word onto a tag, one token at a time. Take it to your word space and find it a home. Tie a thread between your word and its nearest neighbor — and when someone else adds a word, argue kindly about where it lives.

My word:

It lives with these words:

Mark with an X roughly where your word lives:



These words live in my word's neighborhood:  
are. And from that map, the machine learned what a word means — only who its neighbors dog, bed near blanket. The map doesn't know Words that are alike live close together: cat near kept just one thing: a map of where words live. Then the machine threw the books away and

## The map of words

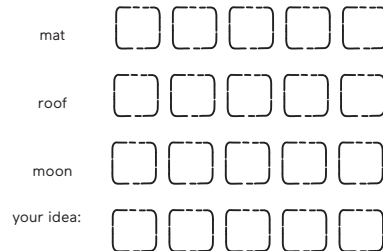
### Task 3:



## Task 6: The machine guesses

The machine never says "this is so." It bets on the word that, in everything it has read, most often followed the words that came before. Usually the language model picks the most likely continuation. Color the squares: how much would you bet on each word?

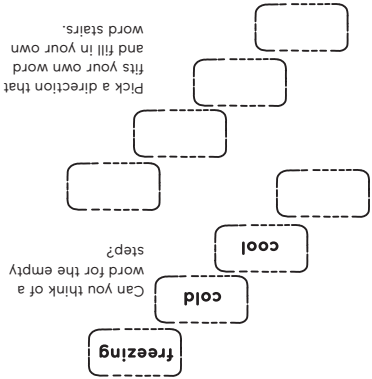
the cat sat on the \_\_\_\_



**Word vector.** To a computer, a word is a string of numbers — an address in a space of thousands of directions. For example, cat's address might be cat = 0.21 -1.03 0.74..., where each number is the word's place on one set of stairs.

**Probability.** A number that tells how certain something is to happen. The sun rising tomorrow: very probable. A dragon knocking on the door: not so much.

Directions — dimensions — come in every kind: fluffiness, shininess, festiveness... The biggest language models have over 16,000 of them! Most don't even have a name, and nobody has explored them all.



## Word stairs

### Task 4:

On the map, every word is also a direction. These words walk in the cold direction:

## Task 5: Calculate with words

On the language model's map, every word has an address made of numbers. And numbers can be calculated with — so words can too!

king — man + woman = queen  
Helsinki — Finland + France = Paris  
cat — small + big = \_\_\_\_\_  
icecream — summer + winter = \_\_\_\_\_

the machine would answer: lion  
· hot chocolate — your answer  
might be better!

My word calculations:

