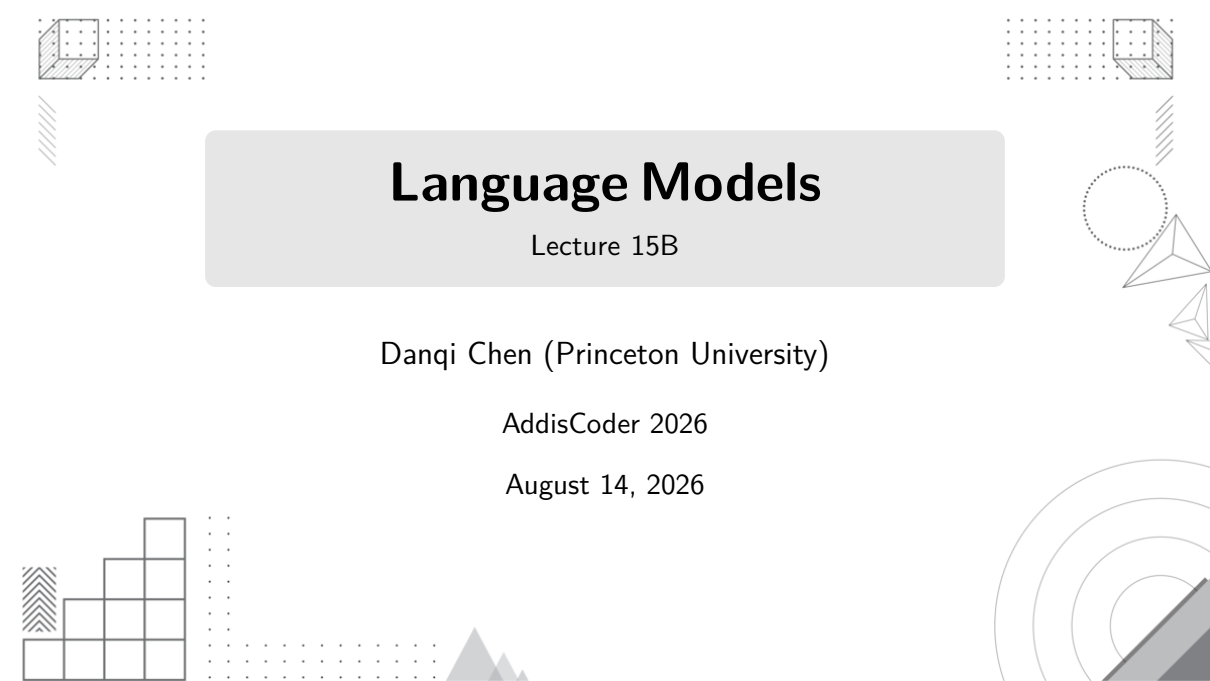




Language Models

Lecture 15B

Danqi Chen (Princeton University)

AddisCoder 2026

August 14, 2026

Tell me about AddisCoder

AddisCoder is a highly intensive, completely free summer computer science program for high-school students from across Ethiopia. It is somewhat unusual because it does **not** just teach students how to code—it introduces them to serious **algorithmic thinking**, often at a level students would normally encounter in university.  AddisCoder +1

It was founded in **2011 by Jelani Nelson**, then a PhD student at MIT and now a professor of computer science at UC Berkeley. The first program was held at the Addis Ababa Institute of Technology with 82 students. Later editions expanded recruitment beyond Addis Ababa, deliberately bringing talented students from public schools and from different regions of Ethiopia.  AddisCoder

What students actually learn

Students start with **Python programming**, and then move surprisingly quickly into topics such as:

- algorithm design and analysis
- asymptotic growth / Big-O ideas
- recursion and recurrences
- induction

Tell me about AddisCoder

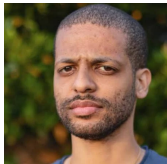
AddisCoder is a highly intensive, completely free summer computer science program for high-school students from across Ethiopia. It is somewhat unusual because it does **not** just teach students how to code—it introduces them to serious **algorithmic thinking**, often at a level students would normally encounter in university.  AddisCoder +1

It was founded in **2011 by Jelani Nelson**, then a PhD student at MIT and now a professor of computer science at UC Berkeley. The first program was held at the Addis Ababa Institute of Technology with 82 students. Later editions expanded recruitment beyond Addis Ababa, deliberately bringing talented students from public schools and from different regions of Ethiopia.  AddisCoder

What students actually learn

Students start with **Python programming**, and then move surprisingly quickly into topics such as:

- algorithm design and analysis
- asymptotic growth / Big-O ideas
- recursion and recurrences
- induction



Jelani Nelson

GPT-2

Released by **OpenAI** in 2019 — an older, much smaller model.

```
The students of AddisCoder are learning how to
```

Generated with <https://huggingface.co/spaces/mkmenta/try-gpt-1-and-gpt-2>

GPT-2

Released by **OpenAI** in 2019 — an older, much smaller model.

The students of AddisCoder are learning how to code in the open source programming language Python.

"We're trying to teach them about the basics of computer science, and we want them to be able to use that knowledge for their own projects," said Addis Coder's founder, Dr. Nabil Al-Khalil. "It's a great way to get people interested in coding."

Al-Khalil is an associate professor at the University of Southern California's School of Cinematic Arts. He has been teaching computer science courses since 2005. ...

Generated with <https://huggingface.co/spaces/mkmenta/try-gpt-1-and-gpt-2>

GPT-2

Released by **OpenAI** in 2019 — an older, much smaller model.

The students of AddisCoder are learning how to code in the open source programming language Python.

"We're trying to teach them about the basics of computer science, and we want them to be able to use that knowledge for their own projects," said Addis Coder's founder, Dr. Nabil Al-Khalil. "It's a great way to get people interested in coding."

Al-Khalil is an associate professor at the University of Southern California's School of Cinematic Arts. He has been teaching computer science courses since 2005. ...

Fluent, confident — and inventing almost all of it.

Generated with <https://huggingface.co/spaces/mkmenta/try-gpt-1-and-gpt-2>

How can a machine possibly know
which word should come next?

How can a machine possibly know
which word should come next?

Today: **concepts only**.

No implementation, no code, no lab.

What is a model?

A **model** is a simplified version of the real world that helps us make **predictions**.

What is a model?

A **model** is a simplified version of the real world that helps us make **predictions**.

- ▶ A **map** is a model of a city. It leaves out almost everything — and still gets you where you are going.

What is a model?

A **model** is a simplified version of the real world that helps us make **predictions**.

- ▶ A **map** is a model of a city. It leaves out almost everything — and still gets you where you are going.
- ▶ A **weather model** looks at **past weather** to predict tomorrow's weather.

What is a model?

A **model** is a simplified version of the real world that helps us make **predictions**.

- ▶ A **map** is a model of a city. It leaves out almost everything — and still gets you where you are going.
- ▶ A **weather model** looks at **past weather** to predict tomorrow's weather.
- ▶ A **language model** looks at **language** and predicts which word is likely to come next.



*“All models are wrong,
but some are useful.”*

— George Box

statistician, 1919–2013

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

The students of AddisCoder are learning how to code in the

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

The students of AddisCoder are learning how to code in the

The students of AddisCoder are learning how to code in the open

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

The students of AddisCoder are learning how to code in the

The students of AddisCoder are learning how to code in the open

The students of AddisCoder are learning how to code in the open source

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

The students of AddisCoder are learning how to code in the

The students of AddisCoder are learning how to code in the open

The students of AddisCoder are learning how to code in the open source

The students of AddisCoder are learning how to code in the open source programming

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

The students of AddisCoder are learning how to code in the

The students of AddisCoder are learning how to code in the open

The students of AddisCoder are learning how to code in the open source

The students of AddisCoder are learning how to code in the open source programming

The students of AddisCoder are learning how to code in the open source programming language

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

The students of AddisCoder are learning how to code in the

The students of AddisCoder are learning how to code in the open

The students of AddisCoder are learning how to code in the open source

The students of AddisCoder are learning how to code in the open source programming

The students of AddisCoder are learning how to code in the open source programming language

⋮

Language models generate one word at a time

A **language model** looks at **language** and predicts which word is likely to come next.

The students of AddisCoder are learning how to code

The students of AddisCoder are learning how to code in

The students of AddisCoder are learning how to code in the

The students of AddisCoder are learning how to code in the open

The students of AddisCoder are learning how to code in the open source

The students of AddisCoder are learning how to code in the open source programming

The students of AddisCoder are learning how to code in the open source programming language

⋮

This is how ChatGPT works. Answering questions, writing code, translating languages, even proving theorems — it is all predicting the next word.

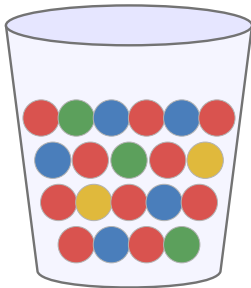
Skittles

Skittles



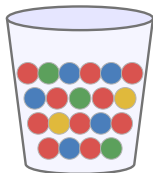
A fruit candy that comes in mixed colours.

A cup of Skittles



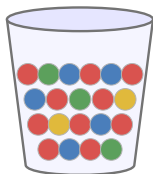
Shake it. Reach in without looking. Take one.

Which colour will I get?



Discuss: Which colour am I most likely to pull out? How much more likely?

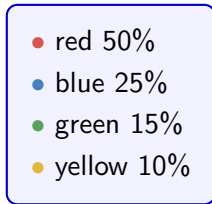
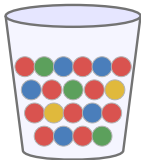
Which colour will I get?



Discuss: Which colour am I most likely to pull out? How much more likely?

• red $\frac{10}{20} = 50\%$ • blue $\frac{5}{20} = 25\%$ • green $\frac{3}{20} = 15\%$ • yellow $\frac{2}{20} = 10\%$

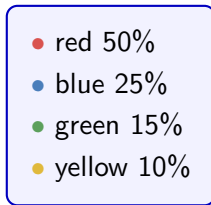
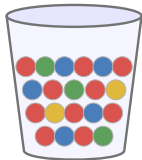
A model of the Skittles cup



all 20, and where each one sits

four numbers

A model of the Skittles cup



all 20, and where each one sits

four numbers

Almost everything thrown away — and it still answers the question.

Count, divide, draw

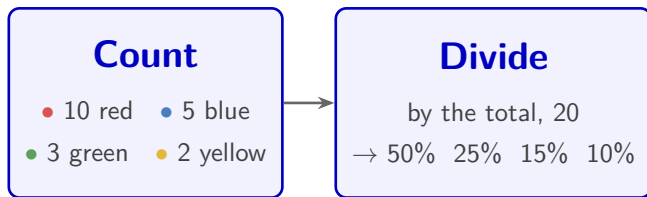
Count, divide, draw

Count

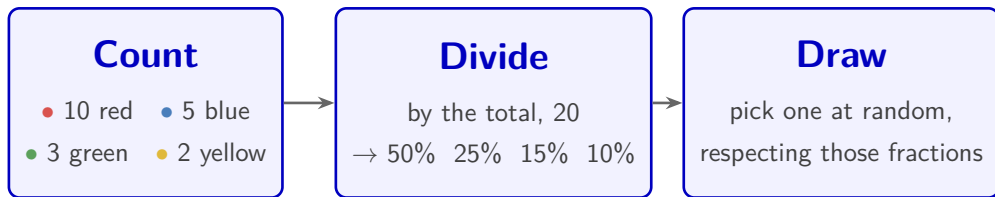
• 10 red • 5 blue

• 3 green • 2 yellow

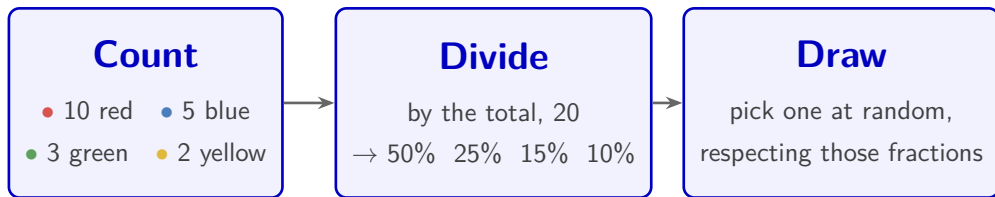
Count, divide, draw



Count, divide, draw



Count, divide, draw



This pattern runs through the whole lecture — only the contents of the cup change.

Fill in the blank

The students opened their _____

Discuss: What word goes in the blank? Shout it out.

What you just did

What you just did

- ▶ You knew which words were **likely** — and which were ridiculous

What you just did

- ▶ You knew which words were **likely** — and which were ridiculous
- ▶ Nobody taught you a rule for this

What you just did

- ▶ You knew which words were **likely** — and which were ridiculous
- ▶ Nobody taught you a rule for this
- ▶ You learned it by **reading and listening**, for years

What you just did

- ▶ You knew which words were **likely** — and which were ridiculous
- ▶ Nobody taught you a rule for this
- ▶ You learned it by **reading and listening**, for years

You are already a language model. Now we build one for a computer.

Today's goal

Give a computer the same ability —
using **count** → **divide** → **draw** and nothing else.

Today's goal

Give a computer the same ability —
using **count** → **divide** → **draw** and nothing else.

The only question is: **what do we put in the cup?**

Attempt 1: put letters in the cup



Count, divide, draw — 60 times in a row.

Attempt 1: the result

```
gaqxbfhavzvevlhdkowlimilhglzkhjxqnhs0 npcoszloskwuomyzcp age
```

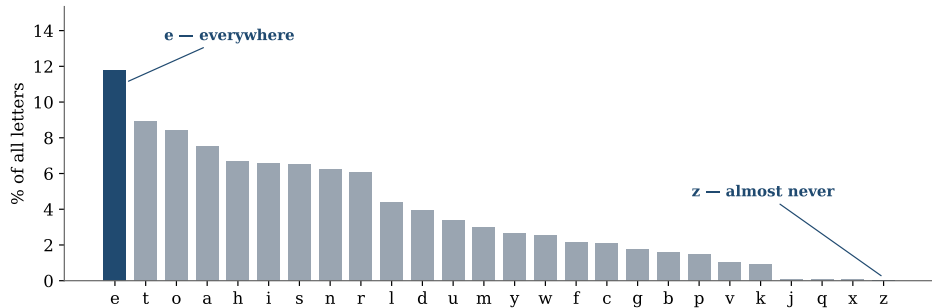
Attempt 1: the result

gaqxbfhazvewlhdkowlimilhglzkhjxqnhs0 npcoszloskwuomyzcp age

Not a single real word.

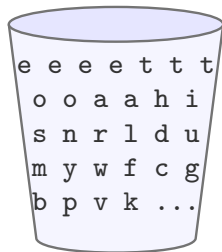
Maybe the ratios are wrong

In real English, e is everywhere and z is rare.
Let's fill the cup to match.



Counted from our Shakespeare text — all 3,297,231 letters of it.

Fill the cup to match



Common letters go in many times, rare ones barely at all —
about **400 e's for every single z.**

Attempt 1, fixed: the result

```
e whontyda nwantf  aap lfnafnogffnha rse  ditt n tir  oragym
```

Attempt 1, fixed: the result

```
e whontyda nwantf  aap lfnafnogffnha rse  ditt n tir  oragym
```

It *looks* like English — until you try to read it.

Words, then

If letters don't work —
put whole **words** in the cup.

Words, then

If letters don't work —
put whole **words** in the cup.

Then every draw is a real word, guaranteed.

Attempt 2: a cup of words



William Shakespeare

Where do we get words?

Shakespeare — his complete works.

826,842	words in total
24,270	different words
41,325	sentences

Attempt 2: a cup of words



William Shakespeare

Where do we get words?

Shakespeare — his complete works.

826,842 words in total
24,270 different words
41,325 sentences

Pour all 826,842 of them
into the cup.

What's in there

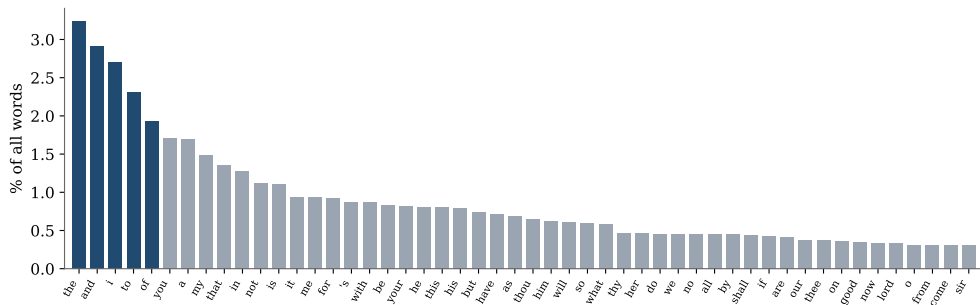
demetrius is a worthy gentleman

so is lysander

i would my father look'd but with my eyes

Real lines from the dataset. Punctuation and capitals already removed.

Count and divide



the 3.2% and 2.9% i 2.7% to 2.3% of 1.9%

Now draw ten words

away count help for look'd you yet brother a but

Now draw ten words

away count help for look'd you yet brother a but

sty varrius my to madam worth tell love hear never

Now draw ten words

away count help for look'd you yet brother a but

sty varrius my to madam worth tell love hear never

is whose which ay patience my bring knowledge for is

Better — and still useless

Better — and still useless

- ▶ **Good:** every single word is a real English word

Better — and still useless

- ▶ **Good:** every single word is a real English word
- ▶ **Good:** common words show up often, rare words rarely

Better — and still useless

- ▶ **Good:** every single word is a real English word
- ▶ **Good:** common words show up often, rare words rarely
- ▶ **Bad:** it is still complete nonsense

Better — and still useless

- ▶ **Good:** every single word is a real English word
- ▶ **Good:** common words show up often, rare words rarely
- ▶ **Bad:** it is still complete nonsense

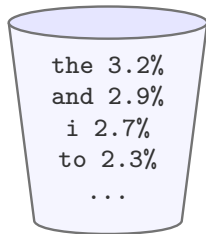
We fixed the spelling. We did not fix the sentence.

Why is it nonsense?

Every draw is completely **independent** of the last one.

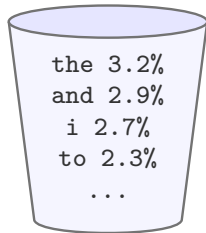
Why is it nonsense?

Every draw is completely **independent** of the last one.



Why is it nonsense?

Every draw is completely **independent** of the last one.



After drawing **the**, the cup is just as happy
to hand you **the** again.

The cup has no memory

It never knows what it just handed you.

The cup has no memory

It never knows what it just handed you.

So let it remember **one** thing:
the word it just handed you.

The cup has no memory

It never knows what it just handed you.

So let it remember **one** thing:
the word it just handed you.

Discuss: How could one cup possibly remember anything?

One cup per word

Don't use one cup. Use **one cup for every word**.



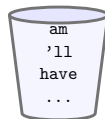
the



to



my



i

One cup per word

Don't use one cup. Use **one cup for every word**.



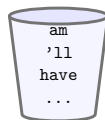
the



to



my



i

A whole **shelf** of cups — one label per word. 24,270 of them.

Look inside a cup

Cup my (12,238 words in it)

lord 13.6%

heart 2.7%

father 2.5%

good 1.9%

life 1.9%

Look inside a cup

Cup my (12,238 words in it)

lord 13.6%

heart 2.7%

father 2.5%

good 1.9%

life 1.9%

Cup good (2,823 words in it)

lord 5.4%

my 4.4%

morrow 4.0%

night 3.7%

sir 2.6%

Look inside a cup

Cup my (12,238 words in it)

lord 13.6%

heart 2.7%

father 2.5%

good 1.9%

life 1.9%

Cup good (2,823 words in it)

lord 5.4%

my 4.4%

morrow 4.0%

night 3.7%

sir 2.6%

*Same three words as before — **count**, **divide**, **draw**. Just once per cup.*

Let's fill the shelf

the cat and the dog sat and talked

Discuss: Which cups do we need? What goes in each one?

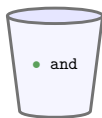
Let's fill the shelf

the cat and the dog sat and talked

Discuss: Which cups do we need? What goes in each one?



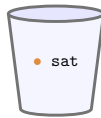
the



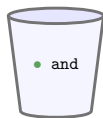
cat



and



dog



sat

Walking from cup to cup



my



good



sooth



i

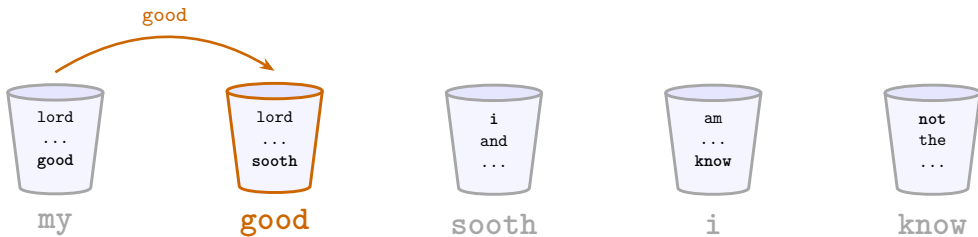


know

my

Draw a word → **walk** to the cup with that label → draw again.

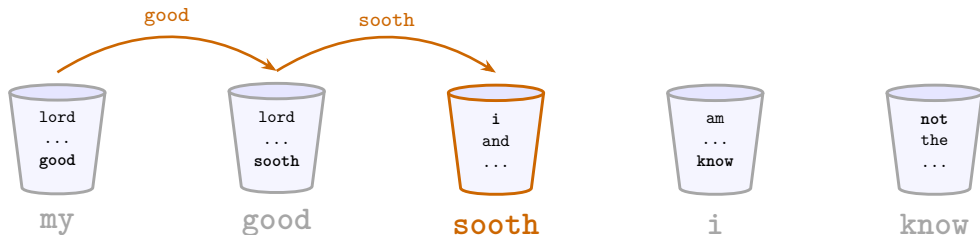
Walking from cup to cup



my good

Draw a word → **walk** to the cup with that label → draw again.

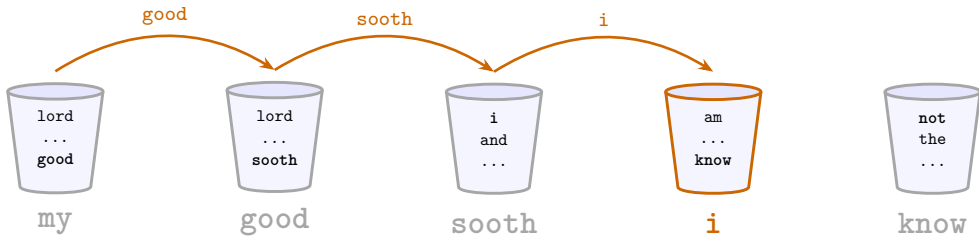
Walking from cup to cup



my good sooth

Draw a word → **walk** to the cup with that label → draw again.

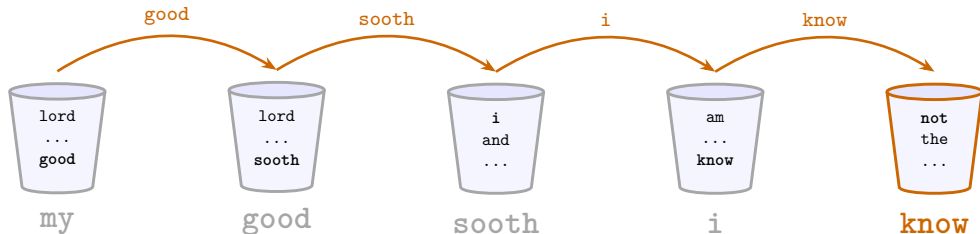
Walking from cup to cup



my good sooth i

Draw a word → **walk** to the cup with that label → draw again.

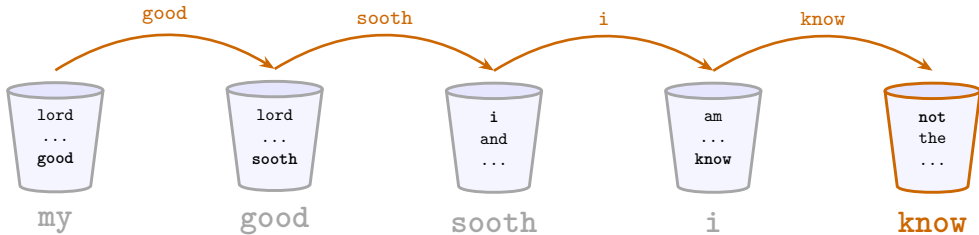
Walking from cup to cup



my good sooth i know

Draw a word → **walk** to the cup with that label → draw again.

Walking from cup to cup



my good sooth i know not

Draw a word → **walk** to the cup with that label → draw again.

Let it walk for twelve words

what is beyond his sword ta'en for both in'an unjust destroy
where 's

Let it walk for twelve words

what is beyond his sword ta'en for both in'an unjust destroy
where 's

what services canst thou speak'st art come near seem'd it
not altogether against

Let it walk for twelve words

what is beyond his sword ta'en for both in'an unjust destroy
where 's

what services canst thou speak'st art come near seem'd it
not altogether against

what thou knowest he means i will not myself there is mis
shapen

One cup, versus one cup per word

One cup (no memory)

is whose which ay patience my bring knowledge for is

One cup, versus one cup per word

One cup (no memory)

is whose which ay patience my bring knowledge for is

One cup per word (one word of memory)

what thou knowest he means i will not myself there is mis
shapen

One cup, versus one cup per word

One cup (no memory)

is whose which ay patience my bring knowledge for is

One cup per word (one word of memory)

what thou knowest he means i will not myself there is mis
shapen

One word of memory. That is the entire difference.

Why stop at one word?

Remember **1** word one cup per word (my $\rightarrow$?)

Remember **2** words one cup per *pair* (my good $\rightarrow$?)

Remember **3** words one cup per *triple* (what is the $\rightarrow$?)

Why stop at one word?

Remember **1** word one cup per word (my $\rightarrow$?)

Remember **2** words one cup per *pair* (my good $\rightarrow$?)

Remember **3** words one cup per *triple* (what is the $\rightarrow$?)

More memory should mean better sentences. Let's check.

Same start, four different memories

Prompt: what is the

No memory

what is the when that between my gone me was and hath i what western

Same start, four different memories

Prompt: what is the

No memory

what is the when that between my gone me was and hath i what western

1 word

what is the oracle thus advises thee call him let me go to say where

Same start, four different memories

Prompt: what is the

No memory

what is the when that between my gone me was and hath i what western

1 word

what is the oracle thus advises thee call him let me go to say where

2 words

what is the letter to him a box o the world and yet it was

Same start, four different memories

Prompt: what is the

No memory

what is the when that between my gone me was and hath i what western

1 word

what is the oracle thus advises thee call him let me go to say where

2 words

what is the letter to him a box o the world and yet it was

3 words

what is the jay more precious than the lark because his feathers are more beautiful

That last one is suspiciously good

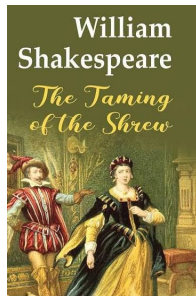
what is the jay more precious than the lark because his
feathers are more beautiful

That last one is suspiciously good

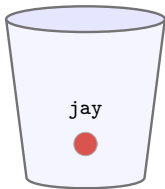
what is the jay more precious than the lark because his
feathers are more beautiful

Every single word of it is
copied straight out of Shakespeare.

(It is a line from *The Taming of the Shrew*.)



Look in the cup



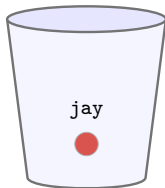
what is the

With **three words** on the label:

570,861 cups on the shelf

703,635 Skittles to share out

Look in the cup



what is the

With **three words** on the label:

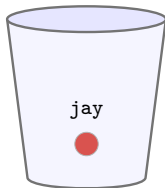
570,861 cups on the shelf

703,635 Skittles to share out

That is **1.2 Skittles per cup**.

91% of them hold exactly one.

Look in the cup



what is the

With **three words** on the label:

570,861 cups on the shelf

703,635 Skittles to share out

That is **1.2 Skittles per cup**.

91% of them hold exactly one.

*A cup with one Skittle isn't a choice. It's a lookup — it isn't writing, it's **reciting**.*

And three words isn't even enough

The students opened their _____

And three words isn't even enough

The students opened their _____

As the proctor started the clock,
the students opened their _____

And three words isn't even enough

The students opened their _____

As the proctor started the clock,
the students opened their _____

To look back **ten** words we would need way, way, way more cups
than we have **Skittles** to put in them.

Modern language models (e.g., GPT-2, ChatGPT)



Modern language models (e.g., GPT-2, ChatGPT)



They don't count cups.

They **merge similar cups together**, using a giant neural network called a **Transformer**.

Modern language models (e.g., GPT-2, ChatGPT)

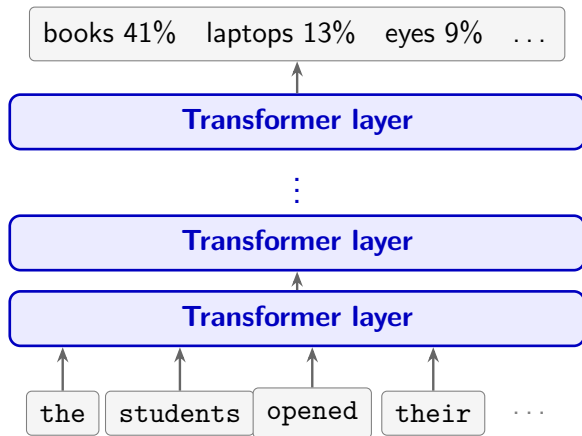


They don't count cups.

They **merge similar cups together**, using a giant neural network called a **Transformer**.

That is how they look back past 1,000 words, even a million words, to draw the next word.

What a Transformer looks like



billions of numbers inside,
learned from huge
amounts of text

Summary

Summary

- ▶ A cup of **Skittles** — a model in four numbers

Summary

- ▶ A cup of **Skittles** — a model in four numbers
- ▶ A cup of **letters** — garbage

Summary

- ▶ A cup of **Skittles** — a model in four numbers
- ▶ A cup of **letters** — garbage
- ▶ A cup of **words** — real words, but no sentences

Summary

- ▶ A cup of **Skittles** — a model in four numbers
- ▶ A cup of **letters** — garbage
- ▶ A cup of **words** — real words, but no sentences
- ▶ **One cup per word** — and it started to sound like English

Summary

- ▶ A cup of **Skittles** — a model in four numbers
- ▶ A cup of **letters** — garbage
- ▶ A cup of **words** — real words, but no sentences
- ▶ **One cup per word** — and it started to sound like English
- ▶ **One cup per three words** — and it started reciting

Summary

- ▶ A cup of **Skittles** — a model in four numbers
- ▶ A cup of **letters** — garbage
- ▶ A cup of **words** — real words, but no sentences
- ▶ **One cup per word** — and it started to sound like English
- ▶ **One cup per three words** — and it started reciting
- ▶ **Merge the cups** — and you have ChatGPT

Summary

- ▶ A cup of **Skittles** — a model in four numbers
- ▶ A cup of **letters** — garbage
- ▶ A cup of **words** — real words, but no sentences
- ▶ **One cup per word** — and it started to sound like English
- ▶ **One cup per three words** — and it started reciting
- ▶ **Merge the cups** — and you have ChatGPT

*Every language model ever built is still answering your question: **what word comes next?***