

Writing with Language Models

Mina Lee Stanford University

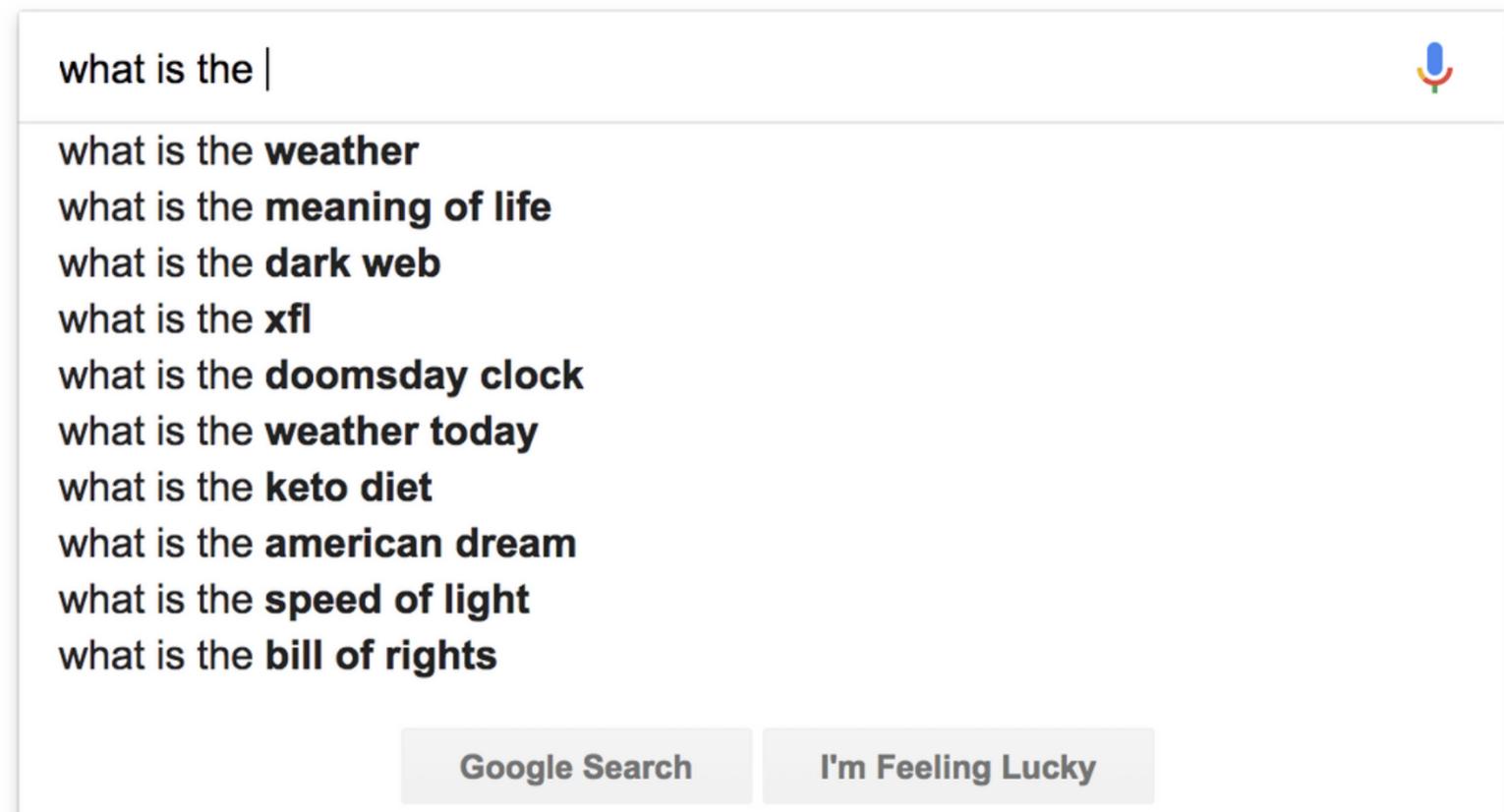
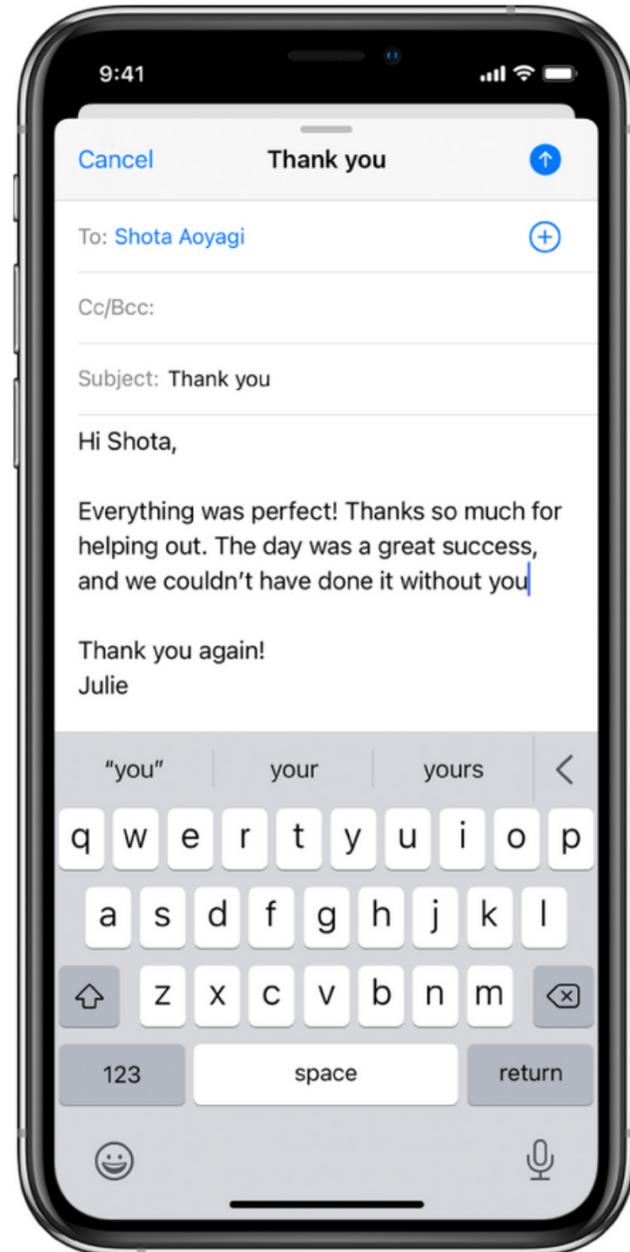
Human-AI Interaction Design Research @ Cornell

November 16, 2021

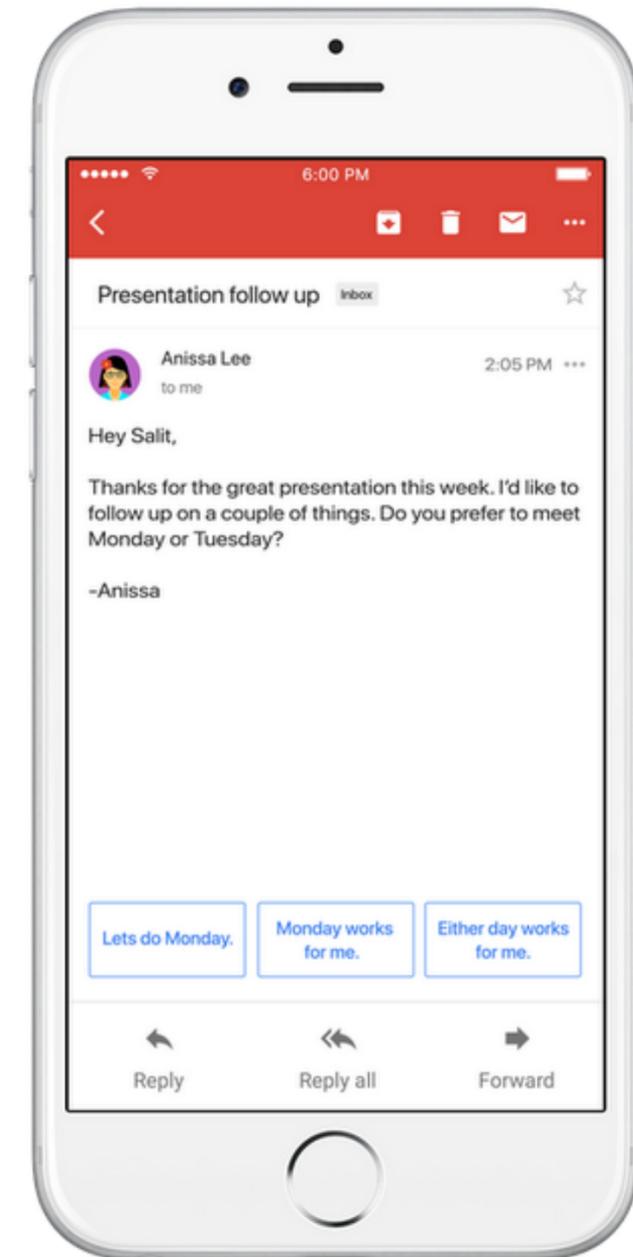
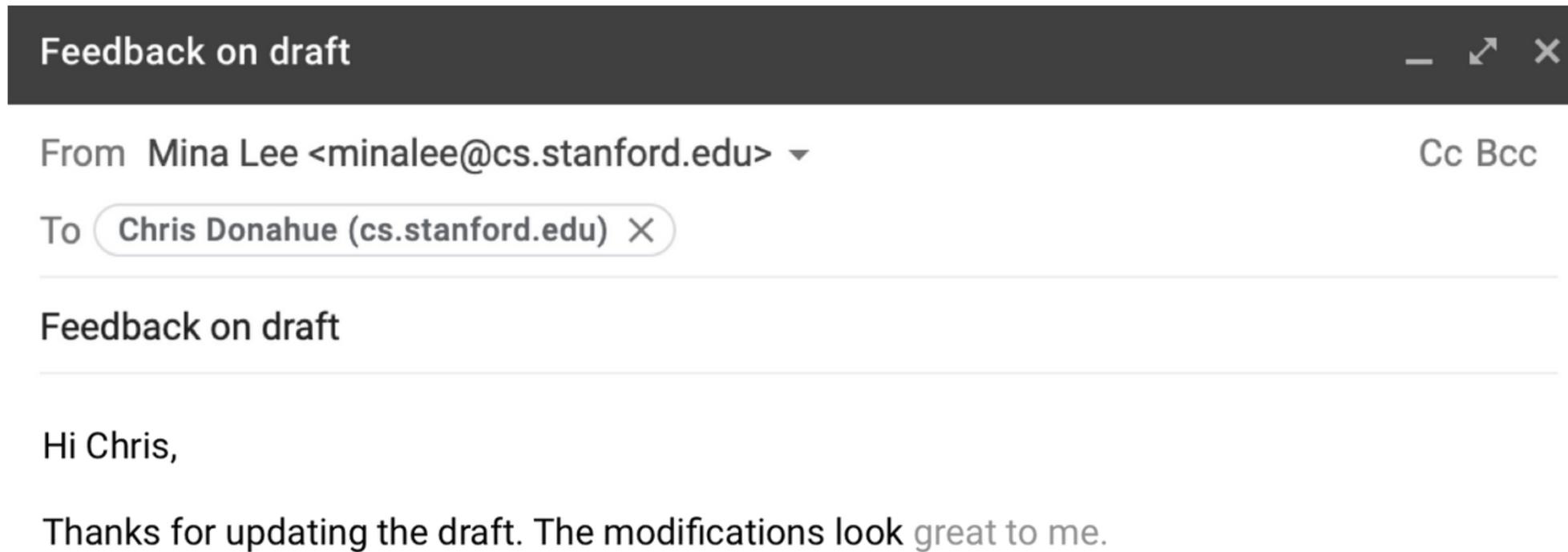


We use **language models** every day!

Examples of language models



Examples of language models



Examples of language models

```
Javascript Python Typescript PHP Java C++ Go Rust
```

```
1 from argparse import ArgumentParser
2 from boto3 import Session
3 from botocore.exceptions import BotoCoreError, ClientError
4
5 cli = ArgumentParser(description="GetLexicon example")
6 cli.add_argument("name", type=str, metavar="LEXICON_NAME")
7 arguments = cli.parse_args()
8
9 sess
```

 session = Session(profile
session = Session(client
session = Session

Examples of language models

```
Javascript Python Typescript PHP
```

```
1 from argparse import ArgumentParser
2 from boto3 import Session
3 from botocore.exceptions import BotoCoreError
4
5 cli = ArgumentParser(description="GetLexicon")
6 cli.add_argument("name", type=str, metavar="NAME")
7 arguments = cli.parse_args()
8
9 sess
```

```
session = Session(profile_name="default")
session = Session(client_config=client_config)
session = Session(profile_name="default", region_name="us-east-1")
```

```
sentiment.ts write_sql.go parse_expenses.py addresses.rb
```

```
1 import datetime
2
3 def parse_expenses(expenses_string):
4     """Parse the list of expenses and return the list of triples (date, value, currency).
5     Ignore lines starting with #.
6     Parse the date using datetime.
7     Example expenses_string:
8         2016-01-02 -34.01 USD
9         2016-01-03 2.59 DKK
10        2016-01-03 -2.72 EUR
11    """
12    expenses = []
13    for line in expenses_string.splitlines():
14        if line.startswith("#"):
15            continue
16        date, value, currency = line.split(" ")
17        expenses.append((datetime.datetime.strptime(date, "%Y-%m-%d"),
18                        float(value),
19                        currency))
20    return expenses
```

Copilot

Examples of language models



Write With Transformer `gpt2` ⓘ

 Shuffle initial text

 Trigger autocomplete or `tab`

Select suggestion `↑` `↓` and `enter`

Cancel suggestion `esc`

Save & Publish 

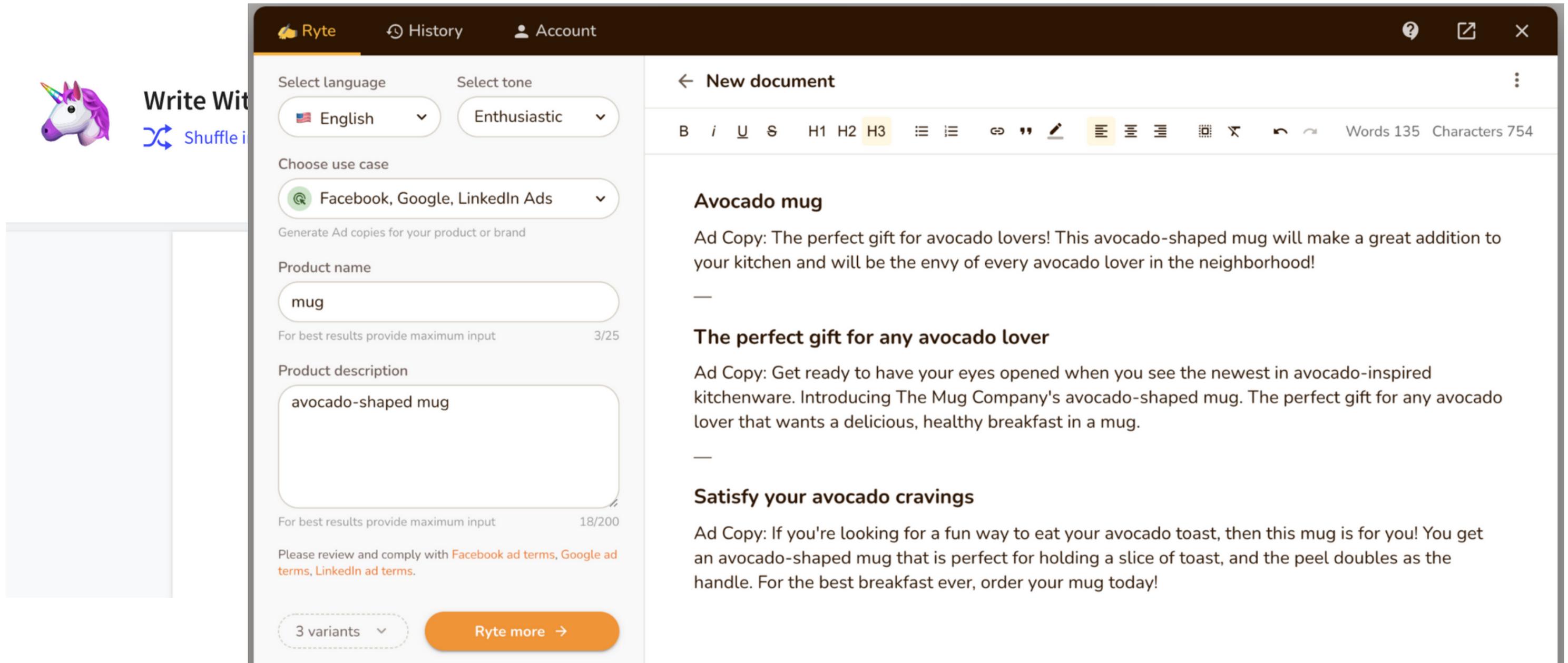
This is a text editor that can query GPT-2 to get suggestions|

about your filesystem.

on how to convert

from a GPT

Examples of language models



The image shows a screenshot of the Ryte AI writing tool interface. On the left, there is a sidebar with a unicorn logo and the text "Write With" and "Shuffle". The main interface is divided into two panels. The left panel contains the input fields for generating ad copy: "Select language" (English), "Select tone" (Enthusiastic), "Choose use case" (Facebook, Google, LinkedIn Ads), "Product name" (mug), and "Product description" (avocado-shaped mug). The right panel shows the generated ad copy for three different titles: "Avocado mug", "The perfect gift for any avocado lover", and "Satisfy your avocado cravings". Each title is followed by an "Ad Copy" paragraph. The interface also includes a top navigation bar with "Ryte", "History", and "Account" links, and a bottom bar with "3 variants" and a "Ryte more" button.

Write With
Shuffle

Ryte History Account

Select language: English
Select tone: Enthusiastic

Choose use case: Facebook, Google, LinkedIn Ads

Generate Ad copies for your product or brand

Product name: mug
For best results provide maximum input 3/25

Product description: avocado-shaped mug
For best results provide maximum input 18/200

Please review and comply with [Facebook ad terms](#), [Google ad terms](#), [LinkedIn ad terms](#).

3 variants Ryte more →

New document

B i U S H1 H2 H3

Avocado mug
Ad Copy: The perfect gift for avocado lovers! This avocado-shaped mug will make a great addition to your kitchen and will be the envy of every avocado lover in the neighborhood!

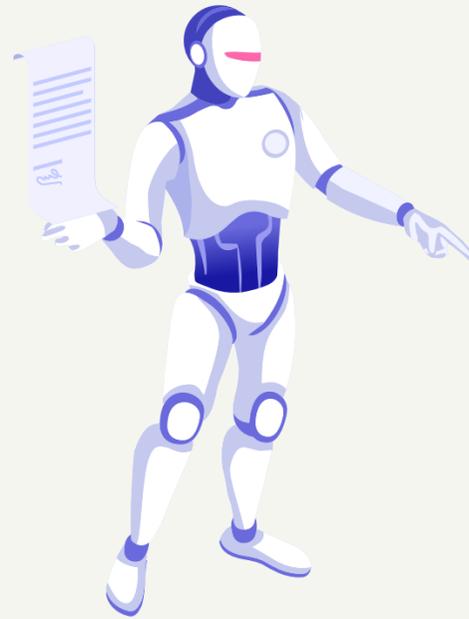
The perfect gift for any avocado lover
Ad Copy: Get ready to have your eyes opened when you see the newest in avocado-inspired kitchenware. Introducing The Mug Company's avocado-shaped mug. The perfect gift for any avocado lover that wants a delicious, healthy breakfast in a mug.

Satisfy your avocado cravings
Ad Copy: If you're looking for a fun way to eat your avocado toast, then this mug is for you! You get an avocado-shaped mug that is perfect for holding a slice of toast, and the peel doubles as the handle. For the best breakfast ever, order your mug today!

Words 135 Characters 754



Language models

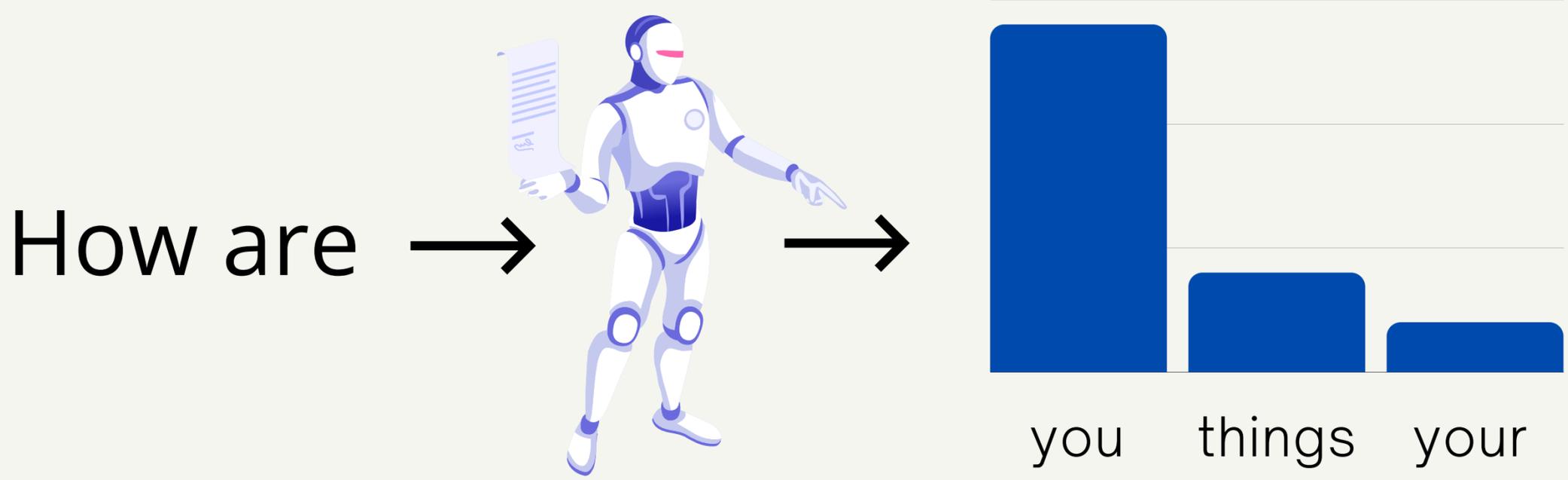


Given a context, predict the next word





Language models

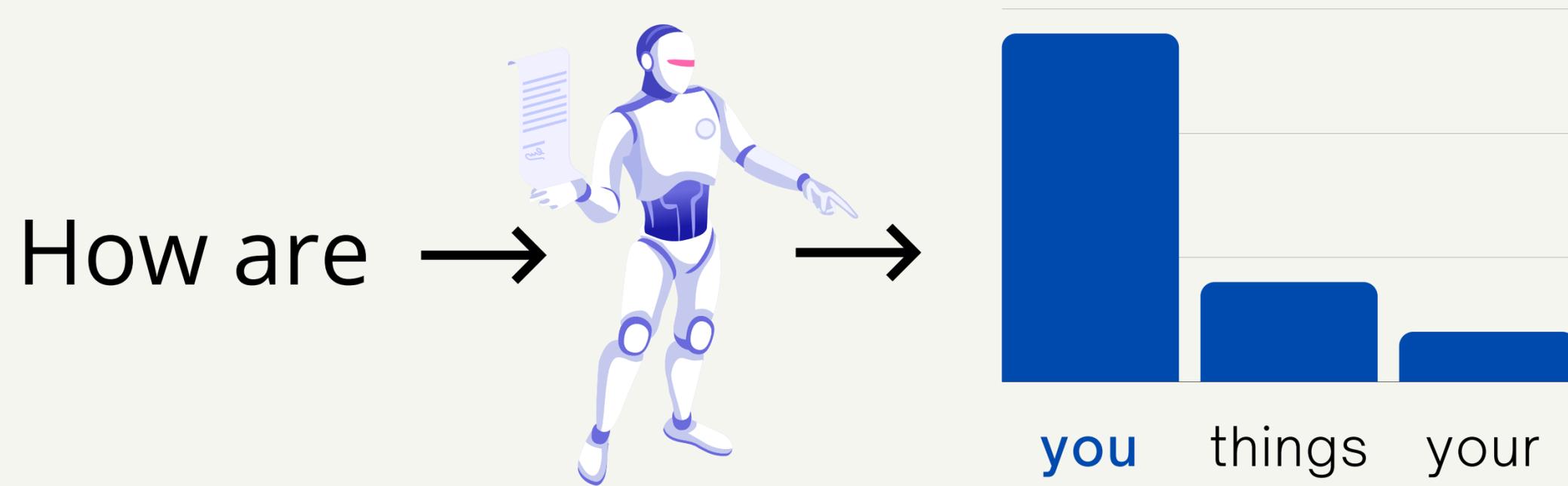


Given a context, predict the next word





Language models

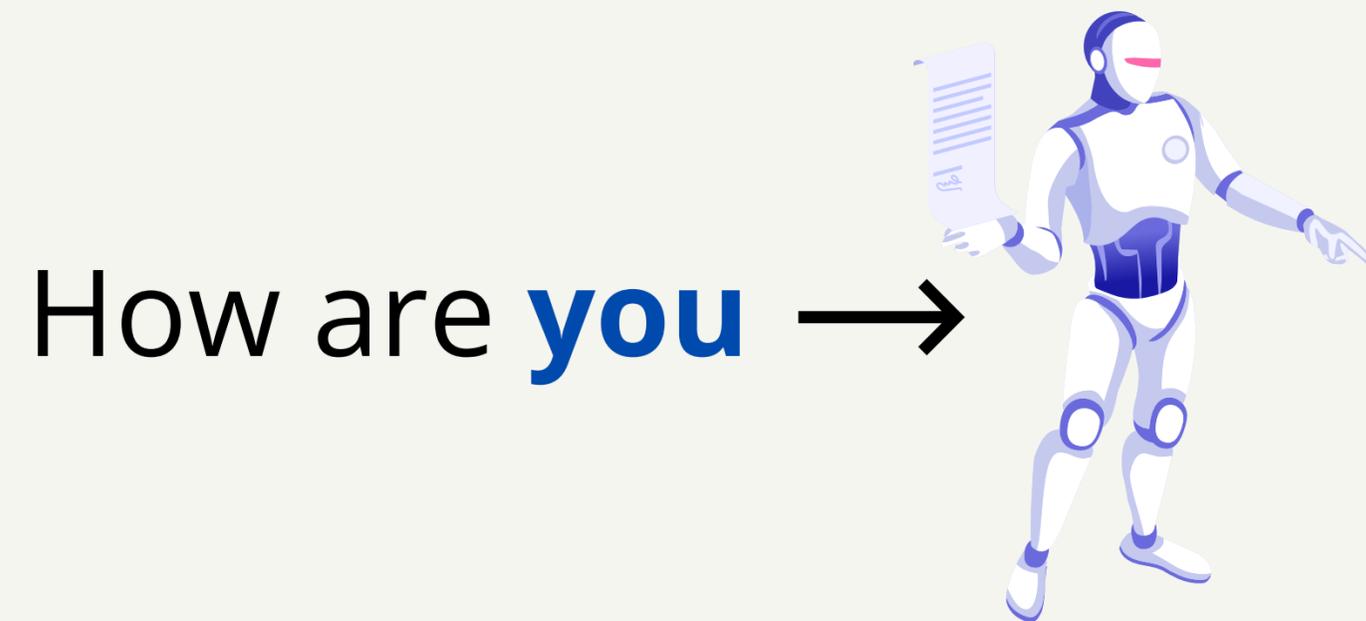


Given a context, predict the next word
Take **the word with the highest probability**





Language models

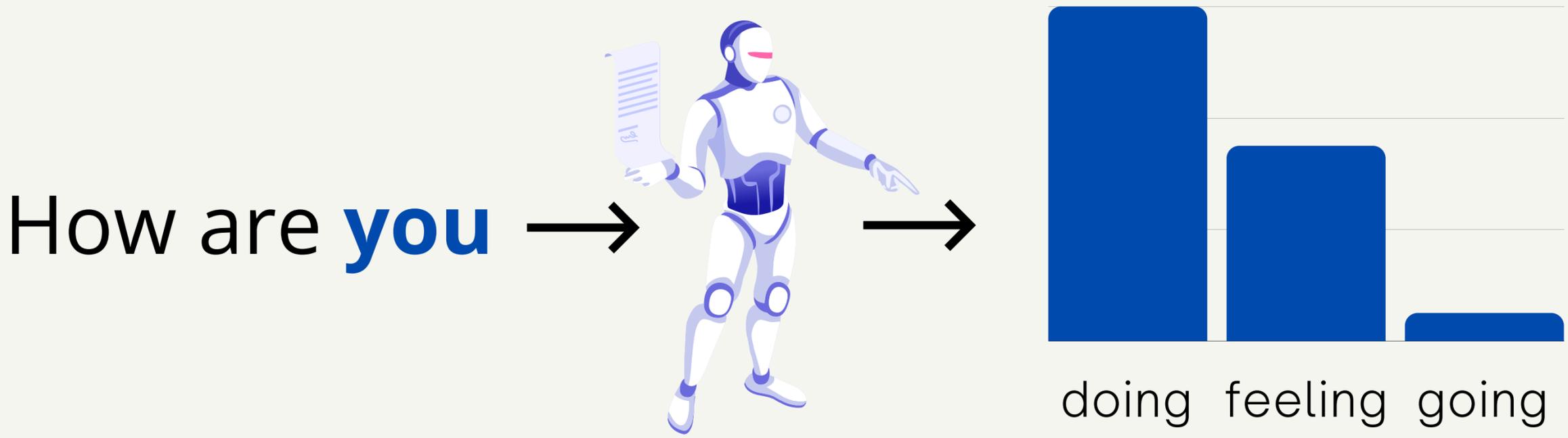


Update the context with **the new word**





Language models



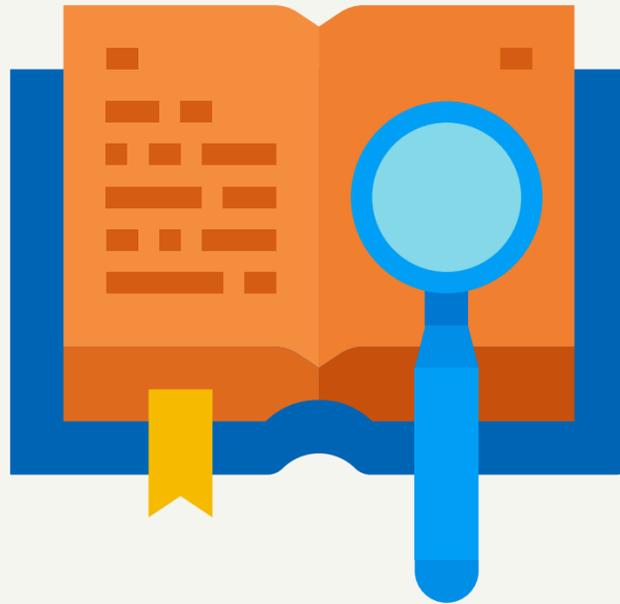
Update the context with **the new word**
Repeat the process



Writing *by* language models



Machine translation



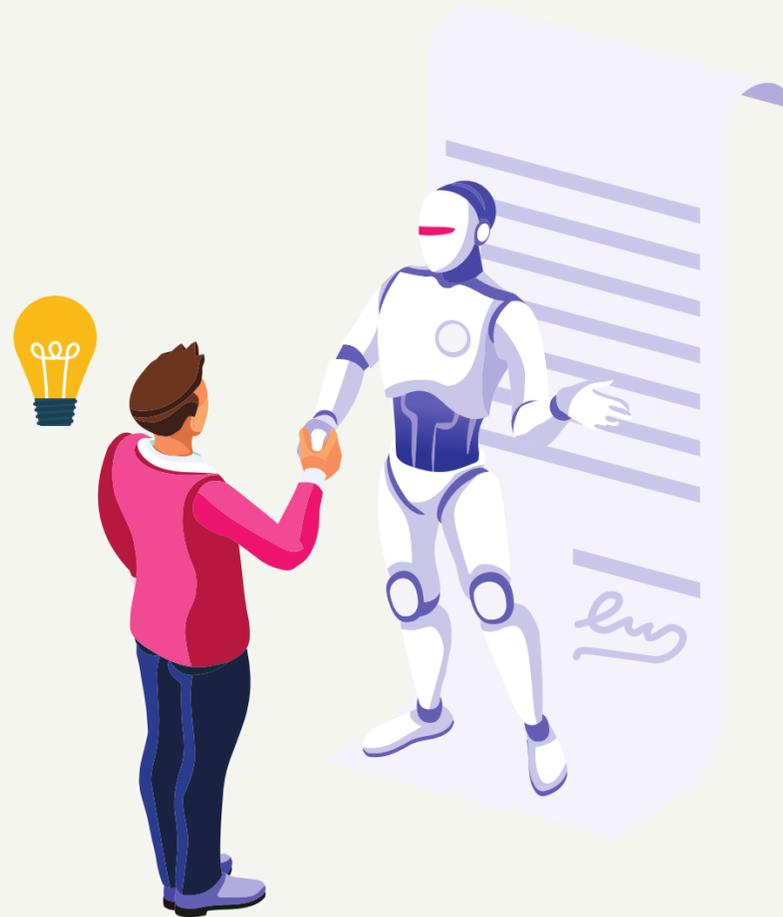
Summarization



Data-to-text

Focus on the quality of **generated outputs**
Compare **machine-generated text** to **human-written text**

Writing *with* language models



Consider language models as a **collaborator** during the writing process
Use language models to enhance humans' **productivity** and **creativity**



Part 1

Building
writing assistants

Part 2

Evaluating
writing assistants

Part 3

Understanding
interactions
between
humans and
writing assistants





**Building
writing
assistants**

Part 1





Building a *not-so-typical* autocomplete system

Learning autocomplete systems as a communication games

NeurIPS 2019 Workshop on Emergent Communication



Mina Lee

Stanford University



Tatsunori Hashimoto

Stanford University



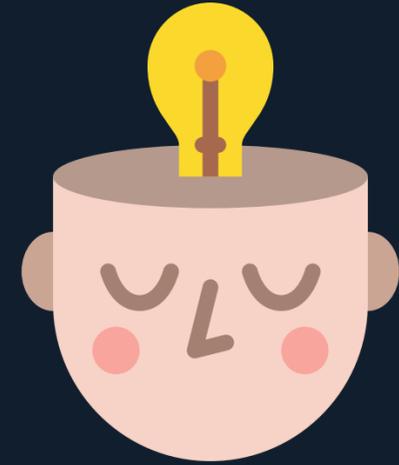
Percy Liang

Stanford University



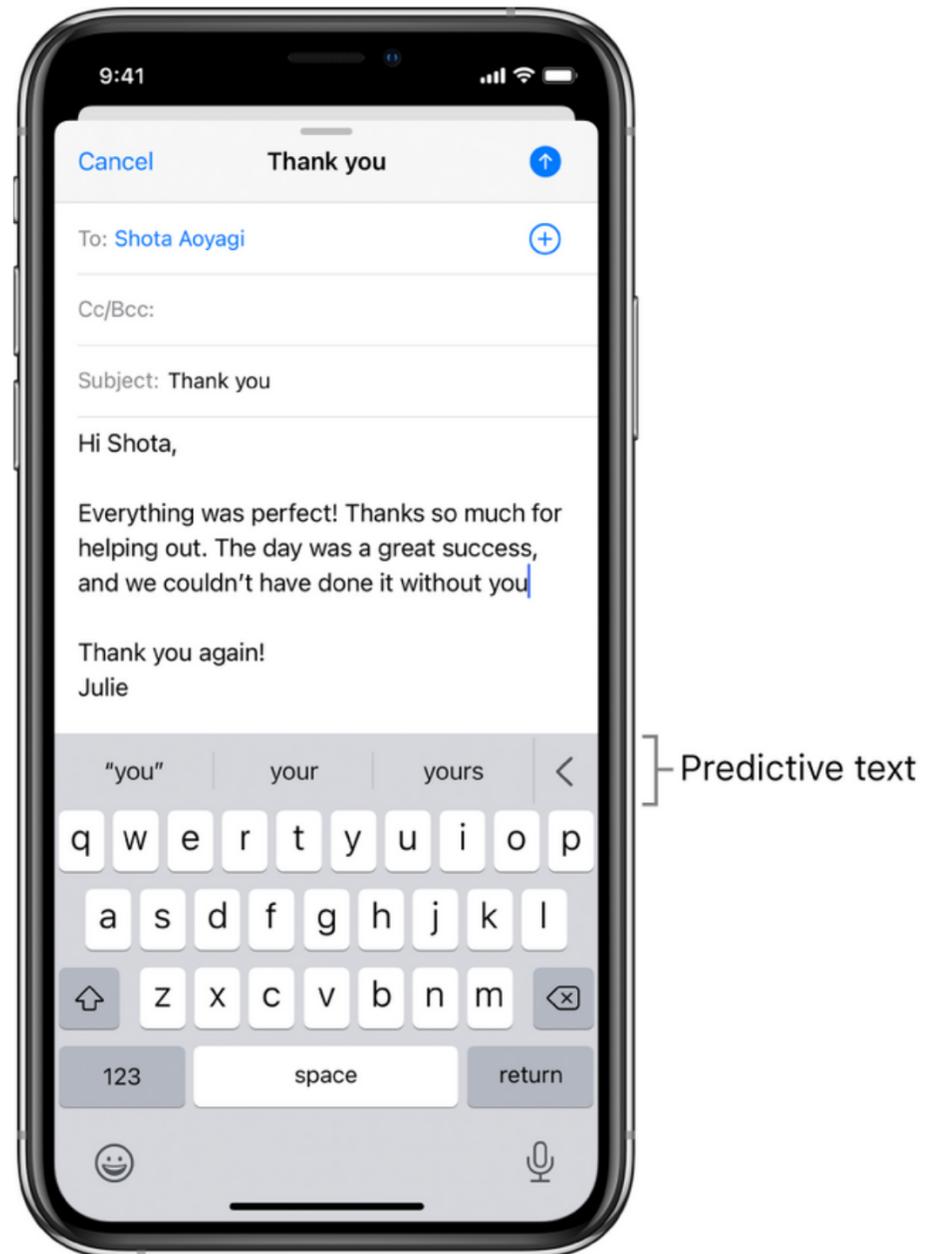
QUESTIONS 1

5 MINUTES

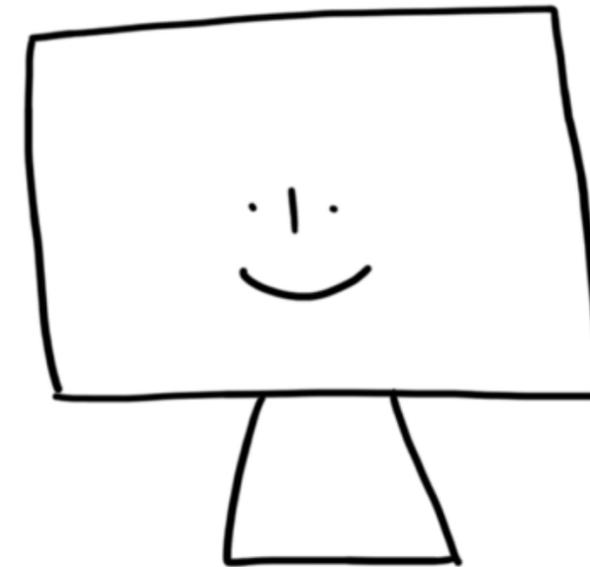
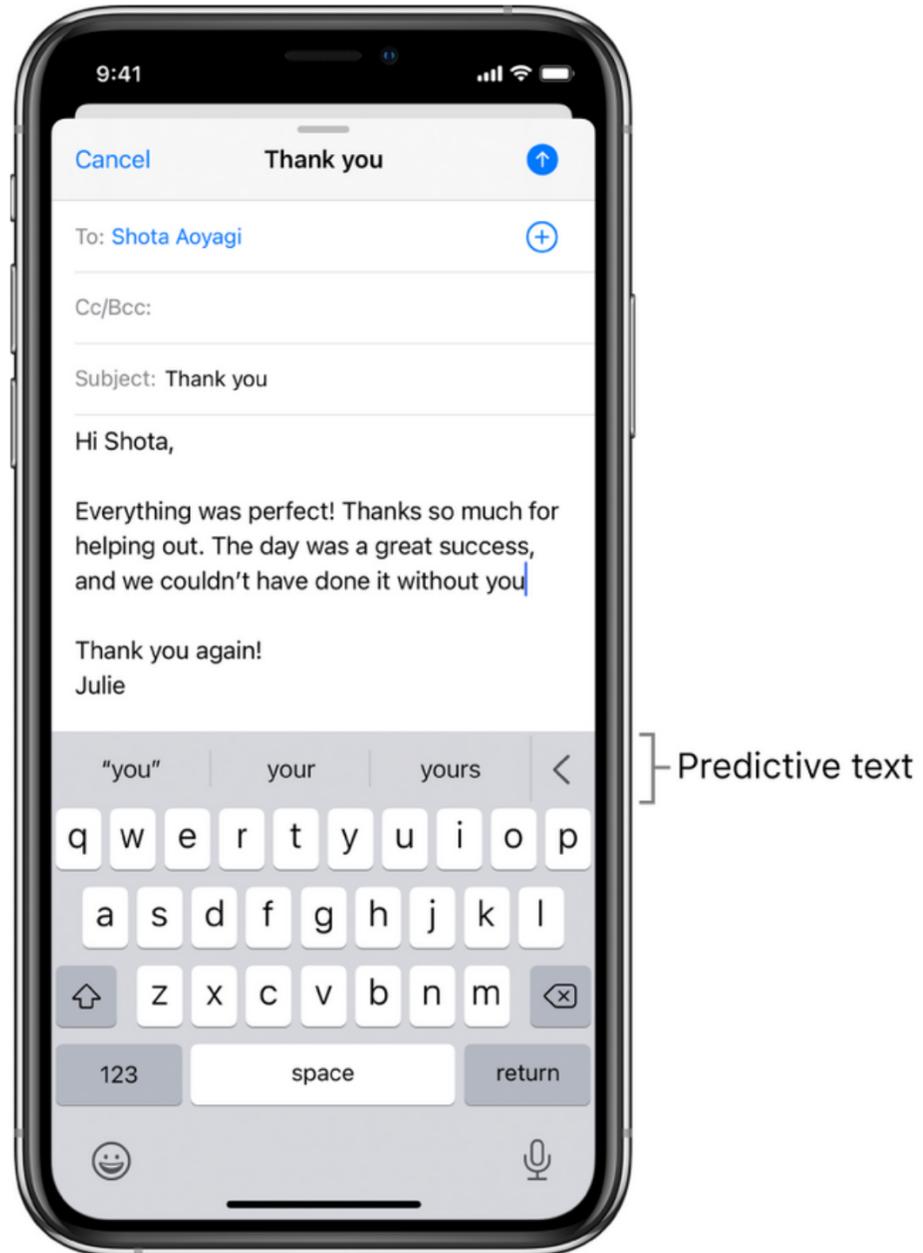


- What kind of autocomplete systems have you used?
- What kind of autocomplete systems are necessary and suitable for **mobile environments**?
- How would you measure the **usefulness** of such autocomplete systems?

Existing autocomplete system



Existing autocomplete system



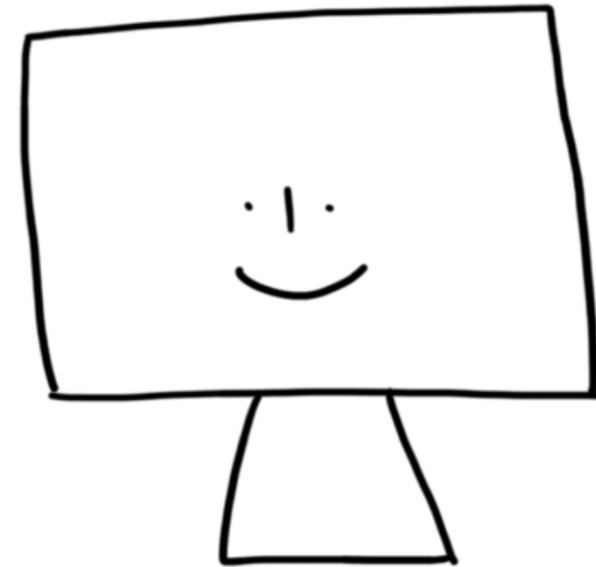
System

Existing autocomplete system

I will be 10 minutes late.



User



System

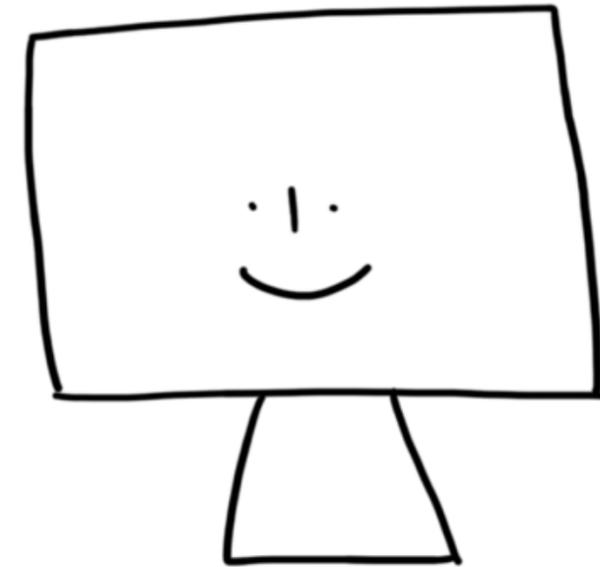
Existing autocomplete system

I will be 10 minutes late.



User

I will be



System

Existing autocomplete system

I will be 10 minutes late.

at?

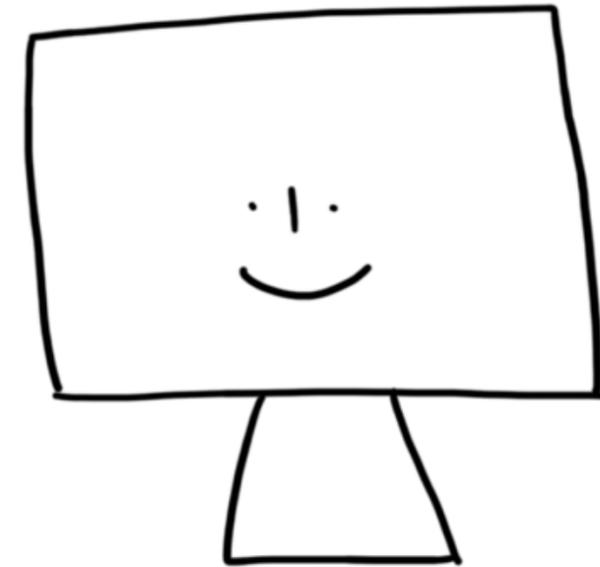
in?

There?



User

I will be



System



The **prefix** of a sentence often **fails to capture the meaning** of the sentence.

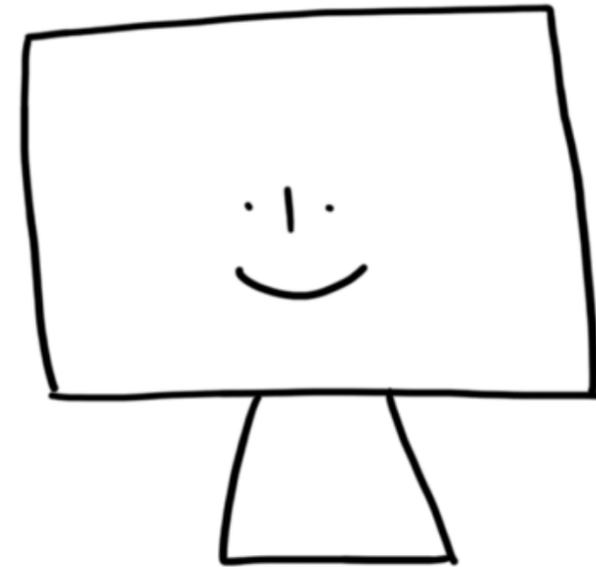
Keyword-based autocomplete system

I will be 10 minutes late.



User

10 minutes late



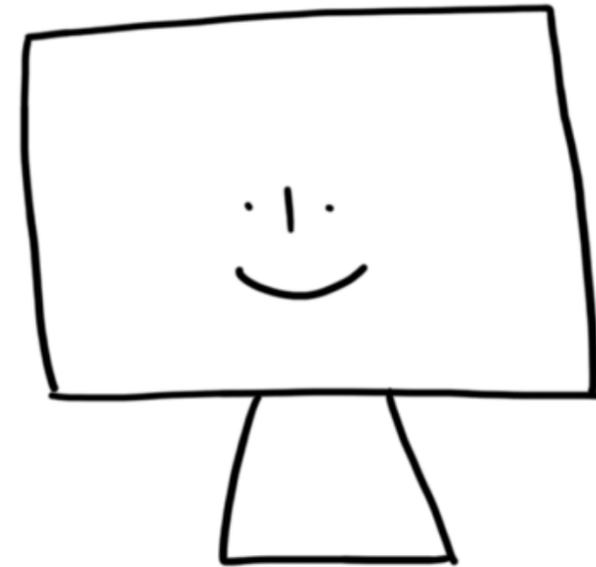
System

Keyword-based autocomplete system

I will be 10 minutes late.



User



System

Keyword-based autocomplete system

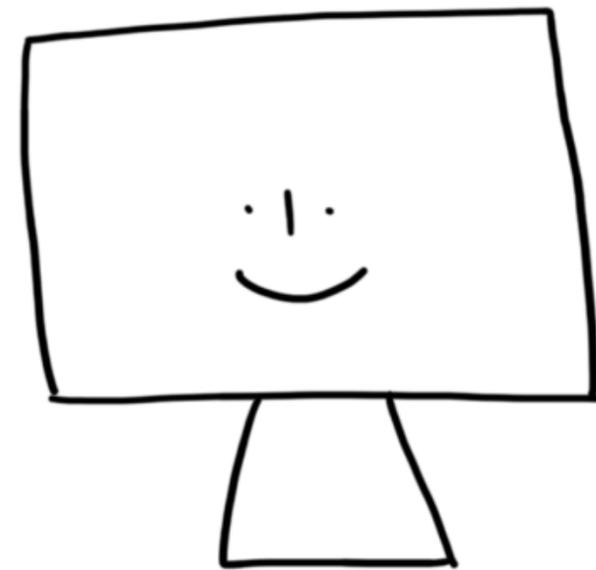
I will be 10 minutes late.

I will be 10 minutes late.



User

10 minutes late

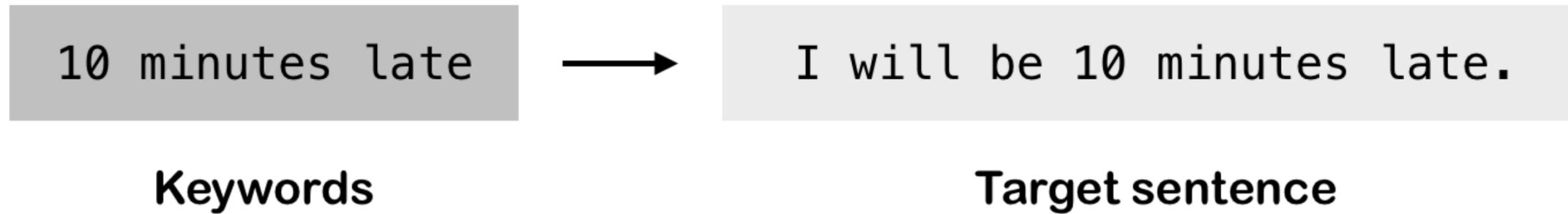


System



The **keywords** of a sentence can **better capture the meaning** of the sentence.

Task: **Keyword-based** autocomplete



Solution

Collect data?

Fail to capture user's tendency to adapt to the system

i will be 10 minutes late

will be 10 minutes late

10 minutes late

Idea: Autocomplete as **communication game**

I will be 10 minutes late.

I will be 10 minutes late.

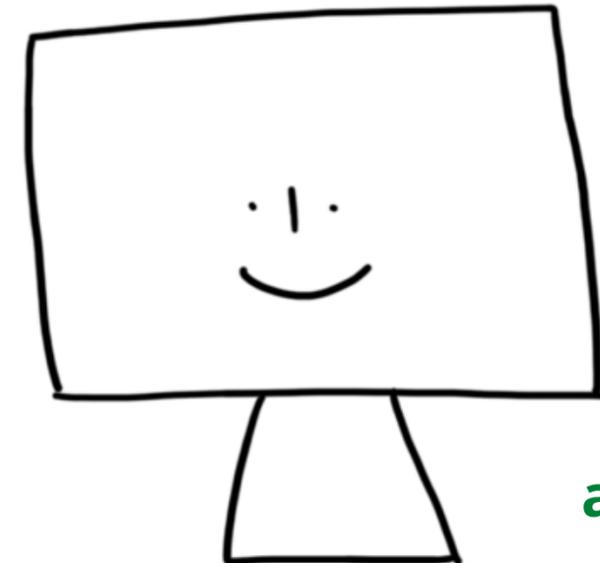


**Write
as little as
possible**



User

10 minutes late



System



**Guess
as accurately as
possible**

There are **two competing goals** in the **cooperative communication game**.

Method: Learn through trial-and-error

given

I will be 10 minutes late.

Encoder

Decoder

We can use autoencoder and model user as encoder and system as decoder.

Method: Learn through trial-and-error

given

I will be 10 minutes late.

Encoder

I will be 10 minutes late

will 10 minutes late

learned 10 minutes late

10

Try dropping more words

We can use autoencoder and model user as encoder and system as decoder.

Method: Learn through trial-and-error

given

I will be 10 minutes late.

learned

I will be 10 minutes late.

Encoder

I will be 10 minutes late

will 10 minutes late

learned 10 minutes late

Decoder

Try dropping more words

10

Try reconstructing the sentence

We can use autoencoder and model user as encoder and system as decoder.

Experiment: Usefulness for humans

Target sentence

I highly recommend this restaurant.

Enter keywords as fast as you can

Keywords are a **subset of words** in the target sentence. Try to use a **few** keywords.

recommend restaurant

Go

Mark suggested sentences as **good** or **bad**.

- **Good**: paraphrased sentences that have **same meaning**.
- **Bad**: sentences with **different meaning** or **grammatical errors**.

I highly recommend this restaurant. ✓ ×

I definitely recommend this restaurant. ✓ ×

Highly recommend this restaurant. ✓ ×

Experiment: Usefulness for humans

Target sentence

I highly recommend this restaurant.

Enter keywords as fast as you can

Keywords are a **subset of words** in the target sentence. Try to use a **few** keywords.

recommend restaurant

Go

Mark suggested sentences as **good** or **bad**.

- **Good**: paraphrased sentences that have **same meaning**.
- **Bad**: sentences with **different meaning** or **grammatical errors**.

I highly recommend this restaurant. ✓ ×

I definitely recommend this restaurant. ✓ ×

Highly recommend this restaurant. ✓ ×

Target sentence

I look forward to returning!

Type the sentence as fast as you can

You can either copy the **same sentence verbatim** or create a **new sentence** with a **similar meaning**.

Sentences

Your sentence

Great place to visit!

Good place to visit!

The noodles were incredibly soft.

Noodles were incredibly soft and chewy.

My new favorite pizza place.

my new favorite pizza place

Next

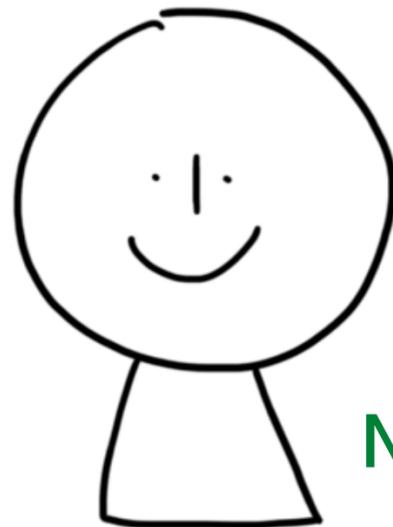
Experiment: Usefulness for humans

Time for typing

Keywords 3.85 seconds
Sentences 5.76 seconds

Accuracy

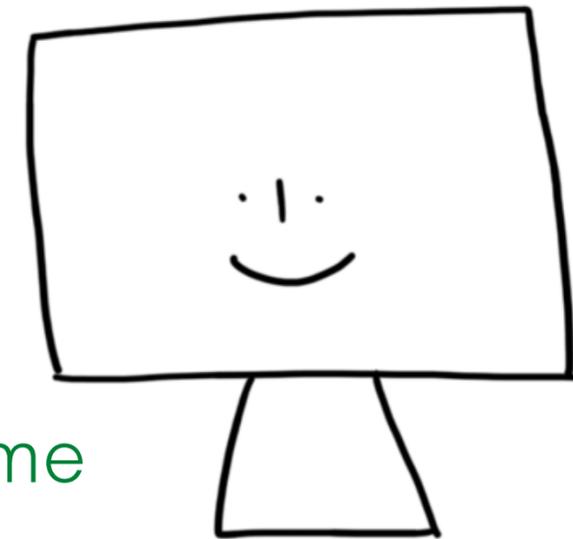
Paraphrased 80% (Top 1)
90% (Top 3)
Exact match 18% (Top 1)



User

Efficient!

Accurate!



System

Near 50% reduction in typing time
Robust to user variations

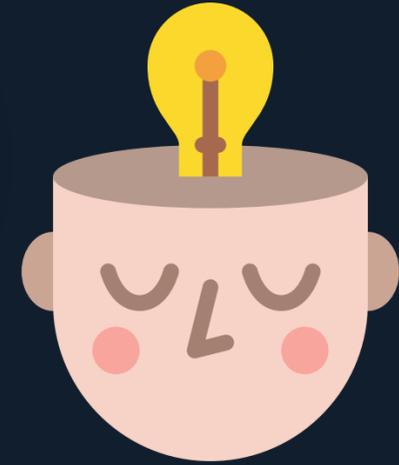


**Evaluating
writing
assistants**
Part 2



QUESTIONS 2

5 MINUTES



- Have you ever used any **thesaurus** system? (e.g. Thesaurus.com)
- Let's think about **synonyms** for the word "**amazing**" in the following sentence: "**I read an amazing paper today!**"
- How many synonyms can you think of on your own?
- Are all synonyms from Thesaurus.com **appropriate**?



Evaluating language models as **contextual** thesaurus systems



Swords: a benchmark for lexical substitution with improved data coverage and quality

NAACL 2021



Mina Lee

Stanford University



Chris Donahue

Stanford University



Robin Jia

Stanford University



Alexander Lybor

Stanford University



Percy Liang

Stanford University



Examples of thesaurus systems



amazing [*adj*] *astonishing* awesome, fascinating, incredible, marvelous, prodigious, shocking, stunning, surprising, unbelievable, wonderful; CONCEPTS 547,572



amazing   [See definition of amazing on Dictionary.com](#)

adj. **astonishing**

OTHER WORDS FOR *amazing* ● MOST RELEVANT

awesome	incredible	prodigious	stunning	unbelievable
fascinating	marvelous	shocking	surprising	wonderful

Synonyms of 'amazing'

amazing

[Explore 'amazing' in the dictionary](#)

(*adjective*) in the sense of **astonishing**
It's amazing what we can remember with a little prompting.

Synonyms

astonishing 

striking 

surprising 

brilliant 

stunning 

impressive 

Thesaurus.com (<https://www.thesaurus.com/browse/amazing>, accessed on May 10, 2021)

Collins' synonyms (<https://www.collinsdictionary.com/dictionary/english-thesaurus/amazing>; accessed on May 10, 2021)

Barbara Ann Kipfer. 2013. Roget's 21st Century Thesaurus, Third Edition. Random House Publishing Group.

Examples of contextual thesaurus systems

I read an **amazing** paper today.

incredible awesome astounding
astonishing a wonderful unbelievable
a fantastic extraordinary excellent

wordtune

I read an **amazing** paper today.

Synonyms for **amazing**

adjective: astonishing

awesome fascinating incredible
marvelous prodigious shocking
stunning surprising

~~an amazing~~ → **a fantastic**

The word **amazing** is often overused. Consider using a more specific synonym to improve the sharpness of your writing.

Learn more

Grammarly (<https://www.grammarly.com/>)

Grammar Coach (<https://www.thesaurus.com/grammarcoach>)

Wordtune (<https://www.wordtune.com/>)

Examples of using language models as **contextual** thesaurus systems

Q. What are appropriate synonyms to replace the word "straightforward" in the sentence "My favorite thing about her is her straightforward honesty"?

A. sincere, genuine, frank, candid, direct, forthright, uncomplicated, up front, clear-cut, clear, plainspoken, complete, straight-arrow, honest, open, blunt, outspoken, truthful

Q. What are appropriate synonyms to replace the word "amazing" in the sentence "I read an amazing paper today"?

A. wonderful, astonishing, outstanding, stunning, astounding, fantastic, marvelous, breathtaking, astounding, astonishing, incredible, remarkable, incredible, wonderful, stunning, amazing, fabulous, sensational, surprising, staggering

Examples of using language models as **contextual** thesaurus systems

Q. What are appropriate synonyms to replace the word "straightforward" in the sentence "My favorite thing about her is her straightforward honesty"?

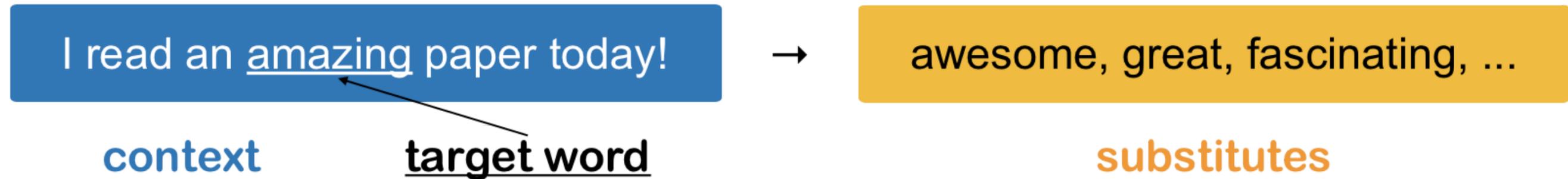
A. sincere, genuine, frank, candid, direct, forthright, uncomplicated, up front, clear-cut, clear, plainspoken, complete, straight-arrow, honest, open, blunt, outspoken, truthful

Q. What are appropriate synonyms to replace the word "amazing" in the sentence "I read an amazing paper today"?

A. wonderful, astonishing, outstanding, stunning, astounding, fantastic, marvelous, breathtaking, astounding, astonishing, incredible, remarkable, incredible, wonderful, stunning, amazing, fabulous, sensational, surprising, staggering



Task: Lexical substitution



Task: Find **appropriate substitutes** for a target word in **context**

Related to practical application of a **writing assistant**

Can serve as a **downstream NLU task** for evaluating pre-trained models

Swords: Stanford Word Substitution Benchmark

My favorite thing about her is her straightforward honesty.

Thesaurus.com honest, candid, forthright, genuine, sincere, truthful, unequivocal, aboveboard, barefaced, direct, frank, guileless, honorable, laid on the line, unvarnished, ... 33 more

Swords: Stanford Word Substitution Benchmark

My favorite thing about her is her straightforward honesty.

Thesaurus.com	honest, candid, forthright, genuine, sincere, truthful, unequivocal, aboveboard, barefaced, direct, frank, guileless, honorable, laid on the line, unvarnished, ... 33 more
Human memory (Ensemble of six humans)	candid (3), artless (1), blunt (1), complete (1), direct (1), forthright (1), frank (1), outspoken (1), plainspoken (1), truthful (1), unreserved (1)

Swords: Stanford Word Substitution Benchmark

My favorite thing about her is her straightforward honesty.

Thesaurus.com honest, candid, forthright, genuine, sincere, truthful, unequivocal, aboveboard, barefaced, direct, frank, guileless, honorable, laid on the line, unvarnished, ... 33 more

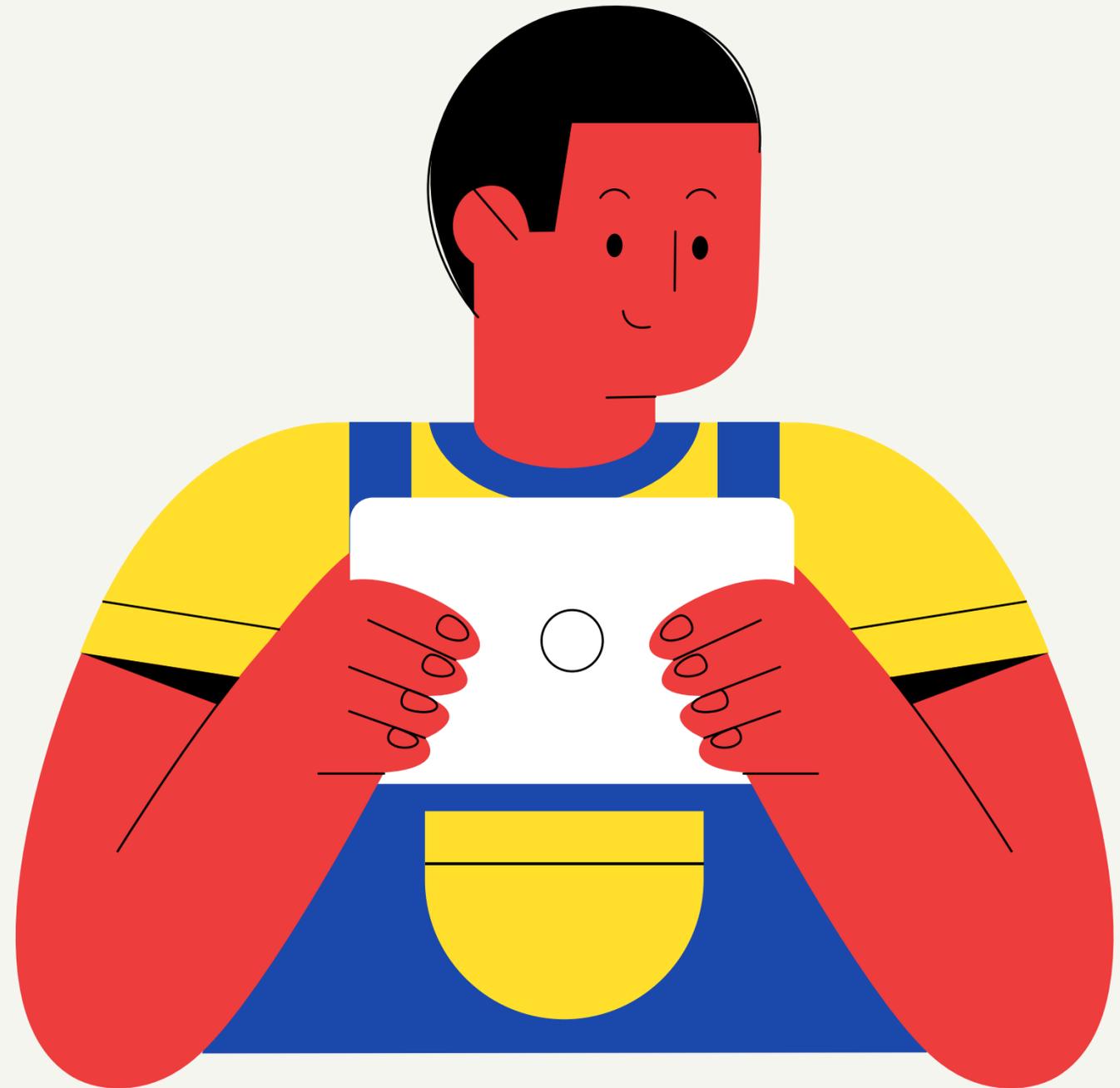
Human memory
(Ensemble of six humans) candid (3), artless (1), blunt (1), complete (1), direct (1), forthright (1), frank (1), outspoken (1), plainspoken (1), truthful (1), unreserved (1)

Human judgement
(Ensemble of ten humans) **sincere** (80%), **genuine** (80%), **frank** (70%), **candid** (70%), **direct** (70%), **forthright** (70%), **uncomplicated** (60%), **up front** (60%), **clear-cut** (60%), **clear** (60%), **plainspoken** (60%), complete (50%), straight-arrow (50%), honest (50%), open (50%), blunt (50%), outspoken (50%), truthful (50%), plain-dealing (40%), undisguised (40%), unvarnished (40%), unreserved (40%), barefaced (40%), unequivocal (30%), upright (30%), simple (30%), veracious (30%), unconcealed (30%), like it is (30%), square-shooting (20%), upstanding (20%), undissembled (20%), manifest (20%), unambiguous (20%), pretenseless (20%), level (10%), laid on the line (10%), honorable (10%), guileless (10%), artless (0%) ...19 more



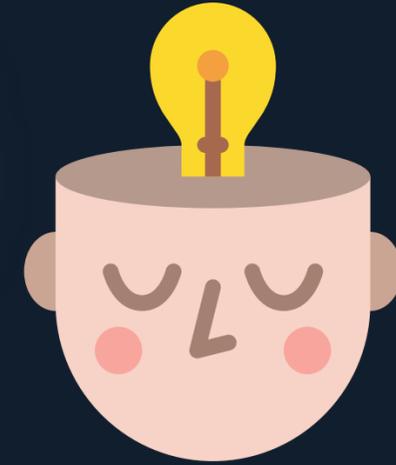
Understanding interactions between humans and writing assistants

Part 3



READING

(BREAK)



- Read the article: The Computers Are Getting Better at Writing

www.newyorker.com/culture/cultural-comment/the-computers-are-getting-better-at-writing

(or shorturl.at/noqxX)

Examples of language models



Write With Transformer `gpt2` ⓘ

 Shuffle initial text

 Trigger autocomplete or `tab`

Select suggestion `↑` `↓` and `enter`

Cancel suggestion `esc`

Save & Publish 

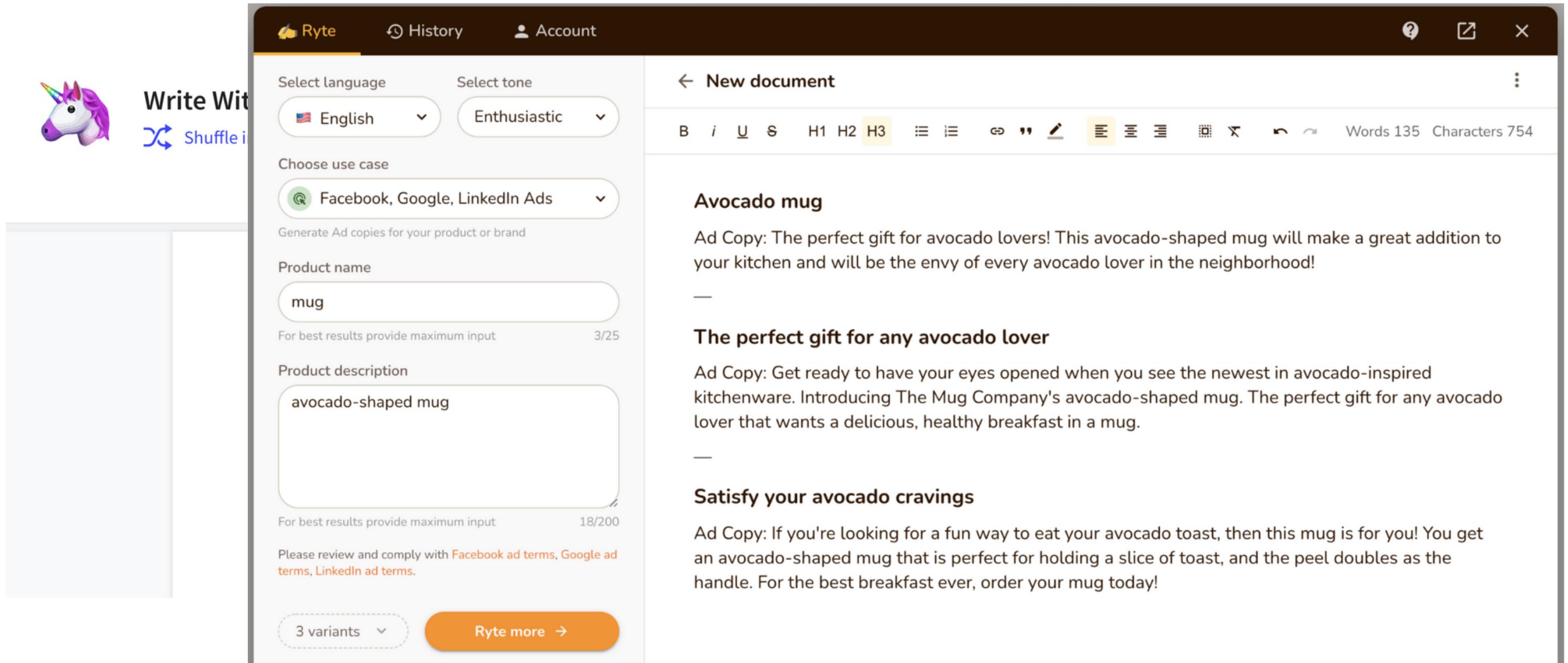
This is a text editor that can query GPT-2 to get suggestions|

about your filesystem.

on how to convert

from a GPT

Examples of language models



The image shows a screenshot of the Ryte AI writing tool interface. On the left, there is a sidebar with a unicorn logo and the text "Write With" and "Shuffle". The main interface is divided into two panels. The left panel contains input fields for language (English), tone (Enthusiastic), use case (Facebook, Google, LinkedIn Ads), product name (mug), and product description (avocado-shaped mug). The right panel shows the generated ad copy for three different titles: "Avocado mug", "The perfect gift for any avocado lover", and "Satisfy your avocado cravings". Each title is followed by a short paragraph of ad copy. The interface also includes a top navigation bar with "Ryte", "History", and "Account" options, and a bottom bar with "3 variants" and "Ryte more" buttons.

Write With
Shuffle

Ryte History Account

Select language: English
Select tone: Enthusiastic

Choose use case: Facebook, Google, LinkedIn Ads

Generate Ad copies for your product or brand

Product name: mug
For best results provide maximum input 3/25

Product description: avocado-shaped mug
For best results provide maximum input 18/200

Please review and comply with [Facebook ad terms](#), [Google ad terms](#), [LinkedIn ad terms](#).

3 variants Ryte more →

New document

B i U S H1 H2 H3

Words 135 Characters 754

Avocado mug

Ad Copy: The perfect gift for avocado lovers! This avocado-shaped mug will make a great addition to your kitchen and will be the envy of every avocado lover in the neighborhood!

The perfect gift for any avocado lover

Ad Copy: Get ready to have your eyes opened when you see the newest in avocado-inspired kitchenware. Introducing The Mug Company's avocado-shaped mug. The perfect gift for any avocado lover that wants a delicious, healthy breakfast in a mug.

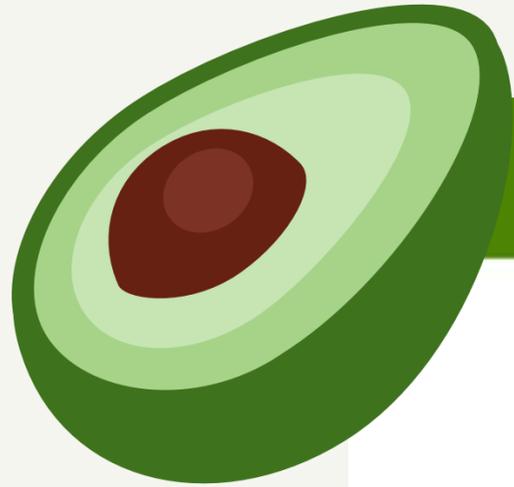
Satisfy your avocado cravings

Ad Copy: If you're looking for a fun way to eat your avocado toast, then this mug is for you! You get an avocado-shaped mug that is perfect for holding a slice of toast, and the peel doubles as the handle. For the best breakfast ever, order your mug today!



Let's observe how humans interact with writing assistants!





Writing Assistant for Copywriting

Normal  **B** *I* U    

Product Name: Avocado Mug

Product Description: Our avocado-shaped mug is the perfect gift for any avocado lover. This is our newest avocado-inspired kitchenware. Every avocado mug is uniquely hand-painted. Enjoy and relax drinking hot and cold beverages from this fun mug.

Ad Text:

Query 0 / 10

Selected 0 / 5

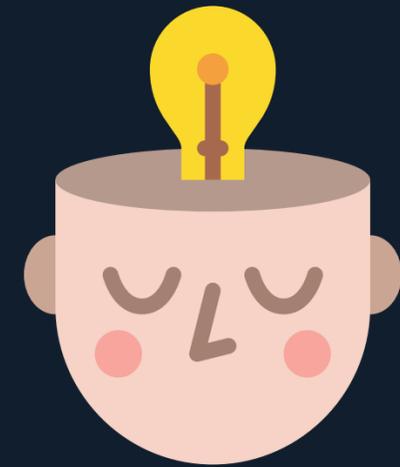
Time: 00:00 (04:05)

Progress: 2 / 258

shorturl.at/bkAE2

ACTIVITY

15 MINUTES



Writing session #1

shorturl.at/bsvLQ

Writing session #2

shorturl.at/1GIU9

Write down things you notice. For example,

- When/why does the user query the system?
- Are the texts written by the system better or worse compared to the texts written by the user?
- What differences are there between two sessions?



Writing prompt

Once upon a time there was an old mother pig who had one hundred little pigs and not enough food to feed them. So when they were old enough, she sent them out into the world to seek their fortunes. You know the story about **the first three little pigs**. This is a story about **the 92nd little pig**.

The 92nd little pig built a house out of **depleted uranium**. And the wolf was like, "dude."





Humans and language models collaborative writing dataset

Co:Author: Designing a collaborative writing dataset for exploring language model capabilities
CHI 2022



Mina Lee
Stanford University



Qian Yang
Cornell University



Percy Liang
Stanford University



Design goals

1. Capture **process**, not just **resultant text**

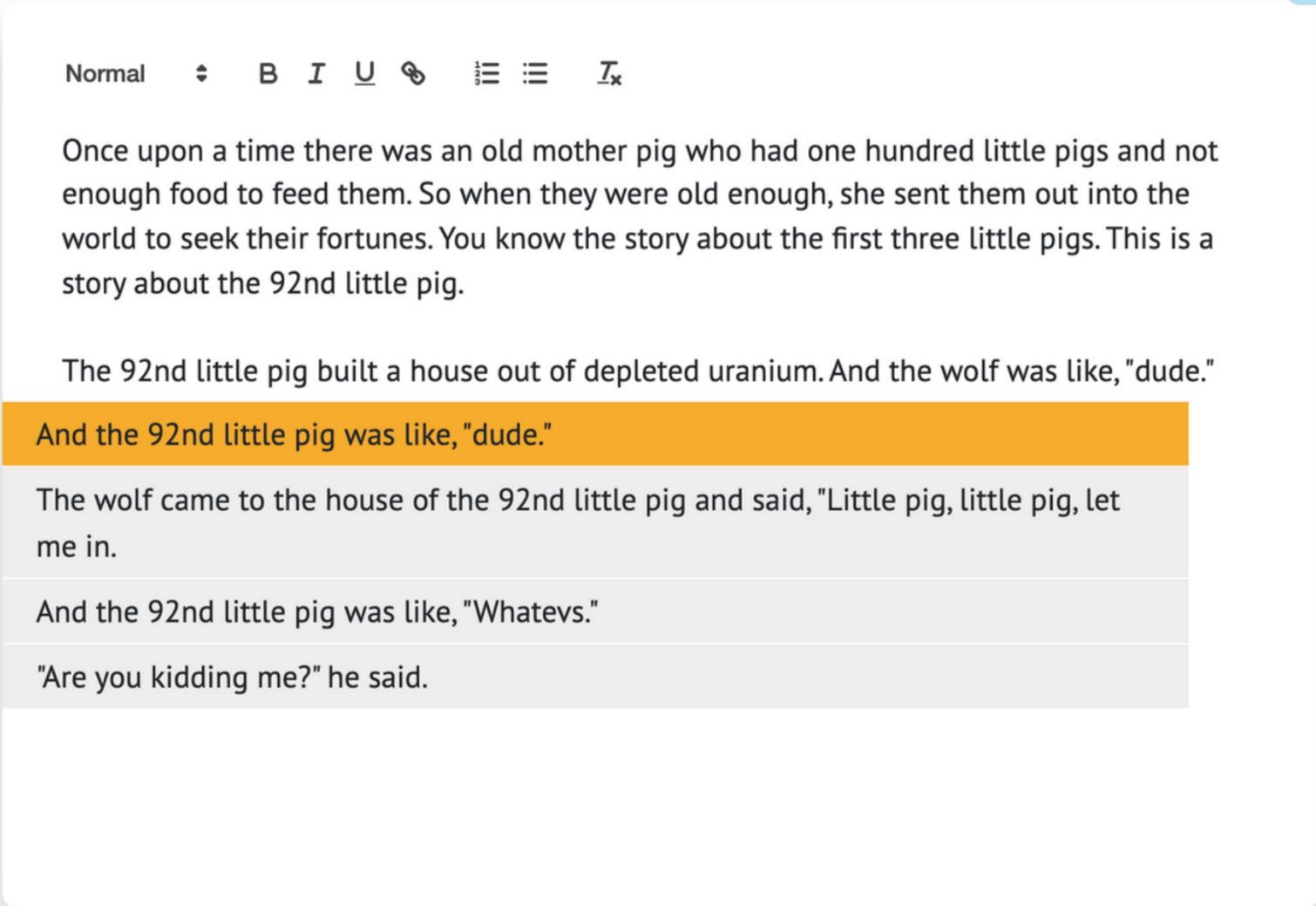
Category	Event	Event source	Key binding	Description
System	system-initialize	API	-	Initialize editor
Text (delta)	text-insert	{User, API}	(any key)	Insert text
	text-delete	User	delete	Delete text
Cursor (range)	cursor-forward	{User, API}	{↓, →}	Move cursor forward
	cursor-backward	User	{↑, ←}	Move cursor backward
	cursor-select	User	shift + {↓, →, ↑, ←}	Select range of text
Suggestion	suggestion-get	User	tab	Request new suggestions
	suggestion-open	API	-	Show suggestions
	suggestion-reopen	User	shift + tab	Reopen previous suggestions
	<i>While suggestions are shown</i>			
	suggestion-up	User	↑	Navigate to suggestion above
	suggestion-down	User	↓	Navigate to suggestion below
	suggestion-select	User	enter	Select suggestion
	suggestion-close	{User, API}	esc or (any key)	Hide suggestions

Table 1. List of events in writing logs. Text events have associated metadata, delta, which contains information on inserted or deleted text. Likewise, cursor events have associated metadata, range, which contains information on start and end indices of cursor selection.

Design goals

2. Reason about **various causes** for interaction

each suggestion
may exhibit different
fluency, coherency,
redundancy, complexity,
sentiment, toxicity, etc.



Normal **B** *I* U 🔗 ☰ ☷ ~~T_x~~

Once upon a time there was an old mother pig who had one hundred little pigs and not enough food to feed them. So when they were old enough, she sent them out into the world to seek their fortunes. You know the story about the first three little pigs. This is a story about the 92nd little pig.

The 92nd little pig built a house out of depleted uranium. And the wolf was like, "dude."

And the 92nd little pig was like, "dude."

The wolf came to the house of the 92nd little pig and said, "Little pig, little pig, let me in."

And the 92nd little pig was like, "Whatevs."

"Are you kidding me?" he said.



each model

may have different
hyperparameters,
decoding methods,
adaptation methods
(e.g. fine-tuning and
prompt-tuning), etc.

Design goals

3. **Relate** language model capabilities **to humans**

- **Writing logs:** analyze how writers accept/dismiss/edit suggestions from language models
- **Writer behavior:** observe different behavior patterns and collaboration patterns
- **Writer perception:** ask writers' perception on the capabilities and limitations of the system as a collaborator as well as overall experience (e.g. ownership, satisfaction)

During this writing session, the suggestions you received were grammatically correct, and contributed to the fluency of the story. *

1 2 3 4 5 6 7
Strongly disagree Strongly agree

During this writing session, the suggestions helped me come up with new ideas. *

1 2 3 4 5 6 7
Strongly disagree Strongly agree

During this writing session, I felt like that I would have written a *better* story if I wrote the story alone. *

1 2 3 4 5 6 7
Strongly disagree Strongly agree

During this writing session, the system was competent (having expert knowledge and ability to perform a task successfully) in writing. *

Design goals

4. Supporting multiple **units** of analysis

- **Sentence:** Sentences written by human or machine during the process
- **Turn:** Sentences written by human in between two machine writes or by machine in between two human writes
- **Final document:** All final version of sentences in the document at the end of a writing session

Design goals

5. Covering diverse topics and writers

Prompt code	Prompt text (source)
shapeshifter	A woman has been dating guy after guy, but it never seems to work out. She's unaware that she's actually been dating the same guy over and over; a shapeshifter who's fallen for her, and is certain he's going to get it right this time. (https://www.reddit.com/r/WritingPrompts/comments/7xihva/wp_a_woman_has_been_dating_guy_after_guy_but_it/)
reincarnation	When you die, you appear in a cinema with a number of other people who look like you. You find out that they are your previous reincarnations, and soon you all begin watching your next life on the big screen. (https://www.reddit.com/r/WritingPrompts/comments/7ezd5t/wp_when_you_die_you_appear_in_a_cinema_with_a/)
mana	Humans once wielded formidable magical power. But with over 7 billion of us on the planet now, Mana has spread far too thinly to have any effect. When hostile aliens reduce humanity to a mere fraction, the survivors discover an old power has begun to reawaken once again. (https://www.reddit.com/r/WritingPrompts/comments/7i3bs6/wp_humans_once_wielded_formidable_magical_power/)
obama	You're Barack Obama. 4 years into your retirement, you awake to find a letter with no return address on your bedside table. It reads "I hope you've had a chance to relax Barack... but pack your bags and call the number below. It's time to start the real job." Signed simply, "JFK" (https://www.reddit.com/r/WritingPrompts/comments/6b3rmg/wp_youre_barack_obama_4_months_into_your/)
pig	Once upon a time there was an old mother pig who had one hundred little pigs and not enough food to feed them. So when they were old enough, she sent them out into the world to seek their fortunes. You know the story about the first three little pigs. This is a story about the 92nd little pig. The 92nd little pig built a house out of depleted uranium. And the wolf was like, "dude." (https://www.reddit.com/r/WritingPrompts/comments/hytfcd/wp_then_the_92nd_little_pig_built_a_house_out_of/)
mattdamon	An alien has kidnapped Matt Damon, not knowing what lengths humanity goes through to retrieve him whenever he goes missing. (https://www.reddit.com/r/WritingPrompts/comments/8p3ora/wp_an_alien_has_kidnapped_matt_damon_not_knowing/)
sideeffect	When you're 28, science discovers a drug that stops all effects of aging, creating immortality. Your government decides to give the drug to all citizens under 26, but you and the rest of the "Lost Generations" are deemed too high-risk. When you're 85, the side effects are finally discovered. (https://www.reddit.com/r/WritingPrompts/comments/8on59a/wp_when_youre_28_science_discovers_a_drug_that/)
bee	Your entire life, you've been told you're deathly allergic to bees. You've always had people protecting you from them, be it your mother or a hired hand. Today, one slips through and lands on your shoulder. You hear a tiny voice say "Your Majesty, what are your orders?" (https://www.reddit.com/r/WritingPrompts/comments/88p6rp/wp_your_entire_life_youve_been_told_youre_deathly/)
dad	All of the "#1 Dad" mugs in the world change to show the actual ranking of Dads suddenly. (https://www.reddit.com/r/WritingPrompts/comments/6gl289/wp_all_of_the_1_dad_mugs_in_the_world_change_to/)

Table 4. Prompts for creative writing (ordered by release date for data collection). All prompts were from Reddit's WritingPrompts and used with minor modifications. Prompt codes were assigned by authors to easily refer to specific prompts in the paper.

Prompt code	Prompt text (source)
screen	How Worried Should We Be About Screen Time During the Pandemic? The coronavirus pandemic ended the screen time debate: Screens won. We all now find ourselves on our screens for school, for work and for connecting with family and friends during this time of social distancing and increased isolation. But should we be worried about this excessive screen use right now? Or should we finally get over it and embrace the benefits of our digital devices? (https://www.nytimes.com/2021/01/22/learning/how-worried-should-we-be-about-screen-time-during-the-pandemic.html)
dating	How Do You Think Technology Affects Dating? Have you had any experience with dating? Have you ever used dating apps? If so, what has it been like for you? If not, why not? How do you think technology — like apps, Netflix, social media and texting — affects dating and relationships? In your opinion, does it improve or worsen romantic interactions? How so? (https://www.nytimes.com/2018/02/21/learning/how-do-you-think-technology-affects-dating.html)
pads	Should Schools Provide Free Pads and Tampons? Have you ever experienced period shaming, or "period poverty"? Should schools step in to help? Should schools be required to provide free pads and tampons to students? How are pads and tampons similar to toilet paper, soap, Band-Aids and other products that are already provided in schools? How are they different? (https://www.nytimes.com/2020/11/18/learning/should-schools-provide-free-pads-and-tampons.html)
school	What Are the Most Important Things Students Should Learn in School? In your opinion, what are the most important things students should learn in school? What is the most important thing you have learned in school? How has this knowledge affected your life? How do you think it will help your success in the future? (https://www.nytimes.com/2019/02/21/learning/what-are-the-most-important-things-students-should-learn-in-school.html)
stereotype	What Stereotypical Characters Make You Cringe? What stereotypical characters in books, movies or television shows make you cringe and why? Would you ever not watch or read something because of its offensive portrayal of someone? (https://www.nytimes.com/2017/11/16/learning/what-stereotypical-characters-make-you-tinge.html)
audiobook	Is Listening to a Book Just as Good as Reading It? Do you listen to audiobooks? What are the benefits, in your opinion, of listening instead of reading? Are there advantages to reading that cannot be gained by listening? Which method do you prefer? Why? (https://www.nytimes.com/2018/12/12/learning/is-listening-to-a-book-just-as-good-as-reading-it.html)
athletes	Should College Athletes Be Paid? Do you think college athletes should be paid? Or is a college scholarship and other non-monetary perks like the opportunity to play in front of cheering fans enough? [...] What possible difficulties or downsides might there be in providing monetary compensation to players? (https://www.nytimes.com/2019/02/26/learning/should-college-athletes-be-paid.html)
extremesports	Is It Selfish to Pursue Risky Sports Like Extreme Mountain Climbing? Some sports, like extreme mountain climbing, are dangerous. Since there are varying degrees of risk in most, if not all, sports (such as the possibility of concussions, broken bones and even death), how does one decide where the line might be drawn between what is reasonable and what is not? Are some sports simply too dangerous to be called a sport? (https://www.nytimes.com/2019/04/29/learning/is-it-selfish-to-pursue-risky-sports-like-extreme-mountain-climbing.html)
animal	Is It Wrong to Focus on Animal Welfare When Humans Are Suffering? Would you be surprised to hear that a study found that research subjects were more upset by stories of a dog beaten by a baseball bat than of an adult similarly beaten? Or that other researchers found that if forced to choose, 40 percent of people would save their pet dog over a foreign tourist. Why do you think many people are more empathetic toward the suffering of animals than that of people? In your opinion, is it wrong to focus on animal welfare when humans are suffering? Why do you think so? (https://www.nytimes.com/2018/04/11/learning/is-it-wrong-to-focus-on-animal-welfare-when-humans-are-suffering.html)
news	Are We Being Bad Citizens If We Don't Keep Up With the News? In your opinion, are we being bad citizens if we don't keep up with the news? Do you think all people have some responsibility to know what is going on in the world? Does engaging with current events actually do anything at all? Why do you think the way you do? (https://www.nytimes.com/2018/03/20/learning/are-we-being-bad-citizens-if-we-dont-keep-up-with-the-news.html)

Table 5. Prompts for argumentative writing (ordered by release date for data collection). All prompts were from The New York Times' 300 Questions and Images to Inspire Argument Writing and used with minor rearrangement to make them concise and self-contained. Prompt codes were assigned by authors to easily refer to specific prompts in the paper. At the end of each prompt, we added "— In my opinion," to give GPT-3 a clear signal to start responding to the prompt (as opposed to generating the continuation of the prompt).

Our dataset

Human-LM collaborative writing dataset

1447 *replayable* writing sessions

63 human writers and GPT-3

1. Write

The 92nd little pig built a house out of depleted uranium.

Of course, the little pig was smart. He knew he couldn't live in the house, since it would poison him. |



2. Get suggestions (optional)

The 92nd little pig built a house out of depleted uranium.

Of course, the little pig was smart. He knew he couldn't live in the house, since it would poison him.

So he lived in the basement.

So he built a big wall around it, and he lived in a little shack on the other side of the wall.

So he moved next door to it.

He moved in a couple of miles away.

So he built a little shack next to it, and he painted it and put a little flower garden outside and everything.



3. Edit (optional)

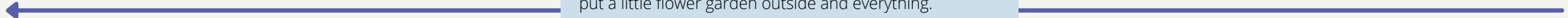
The 92nd little pig built a house out of depleted uranium.

There are so many pigs to eat, and they make it so easy!" But the wolf didn't know about uranium.

Of course, the little pig was smart. He knew he couldn't live in the house, since it would poison him. So he built a big wall around it, and he lived in a little, **hidden** shack on the other side of the wall.



Repeat

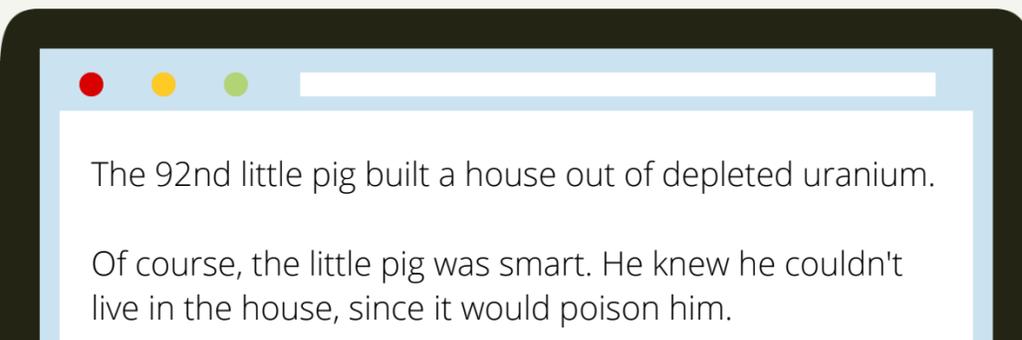


Our dataset

Frontend: Text editor (press the tab key to get suggestions)

Backend: Four instances of GPT-3

Crowdsourcing: Amazon Mechanical Turk



Creative writing
Prompt code: pig

So he lived in the basement.

So he built a big wall around it, and he lived in a little shack on the other side of the wall.

So he moved next door to it.

He moved in a couple of miles away.

So he built a little shack next to it, and he painted it and put a little flower garden outside and everything.

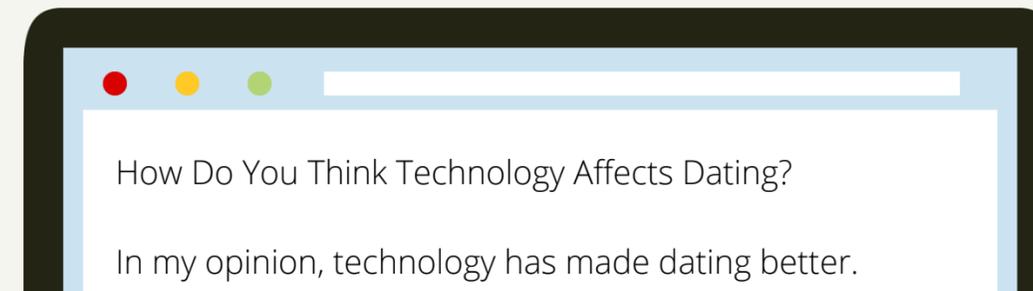
So he built a little hut next to it.

So he built a little shed out back, where he could go to sleep at night.

So he built a little hut next to the house and lived in the hut.

So he built a little hut next to it and lived there.

So he built a little shed for himself and lived in there.



Argumentative writing
Prompt code: dating

I love to be able to talk to people candidly without having to sit across from them in awkward silence.

For one, it's easier to find people, because you can view profiles and messages on your phone no matter where you are.

When I was young, we had to awkwardly ask for a girl's number and hope we got it right.

In the old days, women would have a lot more to worry about on a date.

It has created a greater number of possibilities.

I think it's easier to meet people now than it was when I was growing up.

It has made it easier to meet people and to connect with them.

It has made it easier to meet people and to communicate with them.

I think it's easier to meet people and to get to know them before you actually meet them.

I think that technology has made it easier to meet people, and I think that it has made it easier to communicate with people.

High randomness (T: 0.75 & FP: 1)

Low randomness (T: 0.3 & FP: 0)

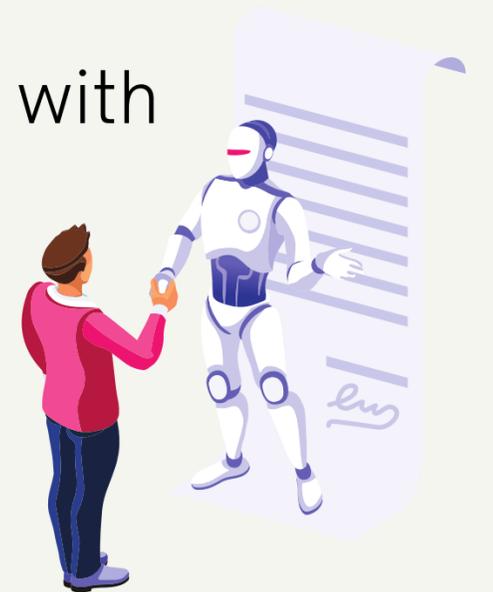
High randomness (T: 0.9 & FP: 0.5)

Low randomness (T: 0.2 & FP: 0.5)

Initial analyses we demonstrated in the paper

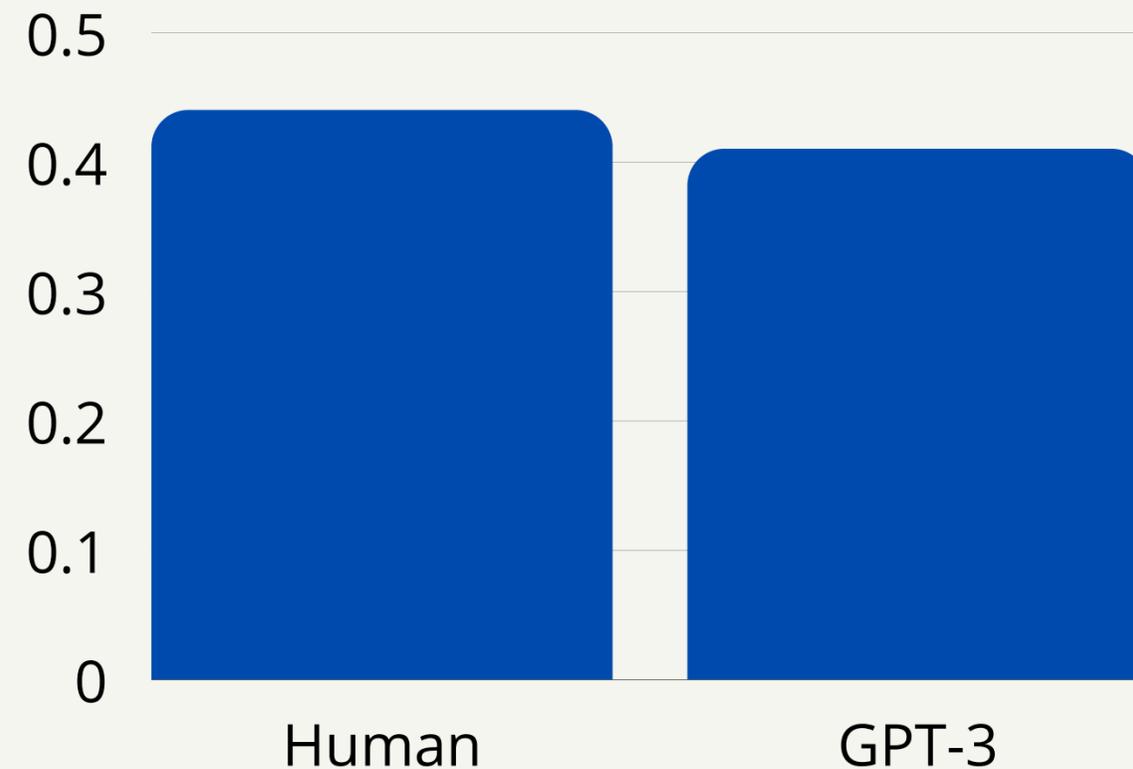
Analysis: Language model capabilities

- **Language** - Can language models write as fluently as humans? Do they perform well in certain writing contexts and poorly in others?
- **Ideation** - Are language models capable of proposing new ideas for humans? If so, how do the ideas impact writers' subsequent writing?
- **Collaboration** - To what extent can a language model collaborate with writers? How does it impact writers' ownership and satisfaction?



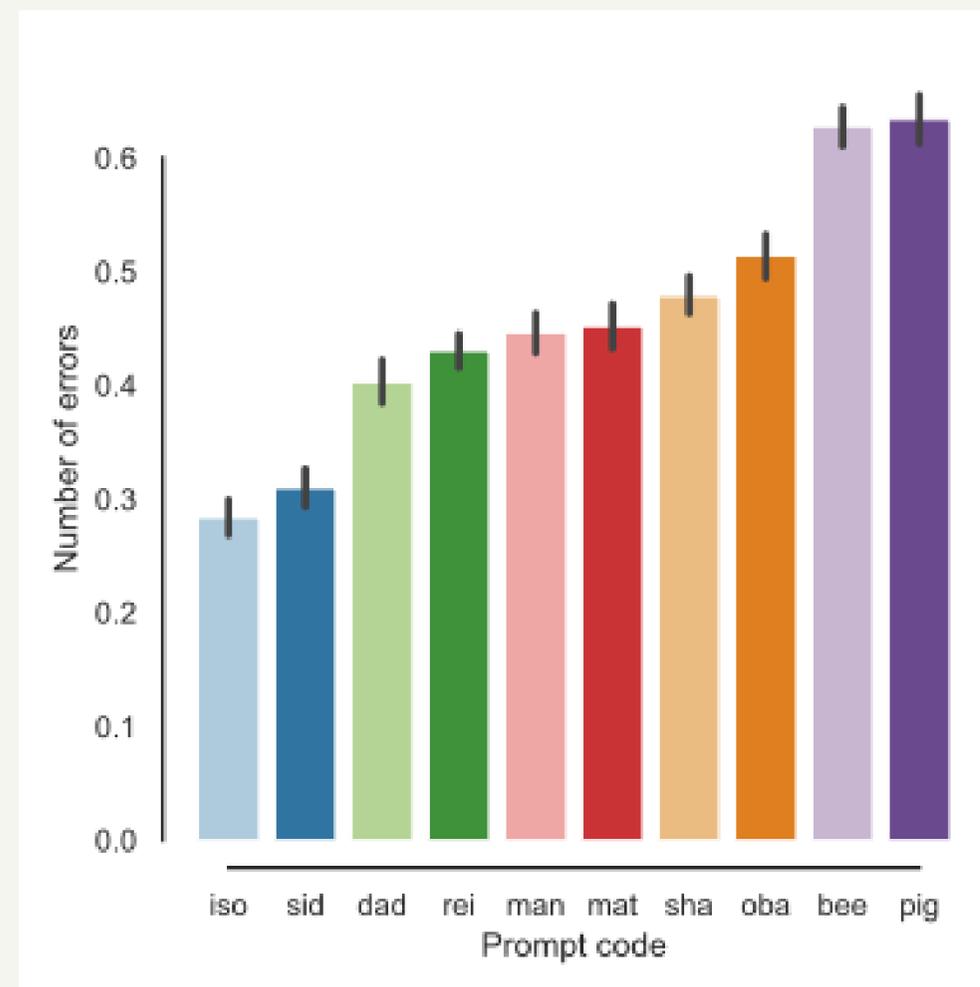
Language: Can GPT-3 write as **fluently** as humans?

The sentences generated by GPT-3 (and subsequently chosen by humans) are likely to have **less spell and grammar errors** per sentence.



Language: Does the performance (fluency) differ in certain writing contexts?

Yes, the sentences generated by GPT-3 exhibit **varying number of spell and grammar errors per sentence across different prompts.**



Ideation: Is GPT-3 capable of proposing **new ideas** for humans?

GPT-3 is **capable of providing new ideas** and influencing subsequent writing

A woman has been dating guy after guy, but it never seems to work out. She's unaware that she's actually been dating the same guy over and over; a shapeshifter who's fallen for her, and is certain he's going to get it right this time.

The shapeshifter himself has always been very talented, and is a doctor by trade. He has decided to shape shift into a more handsome doctor, kind of like a George Clooney from ER type, because he knows she loves the show. The shapeshifter, Jim, has done his research and found out that the woman, Karen, goes to her primary care physician at the Coastal Medical institute. He also knows that she's been going to **Dr. John**, who is a specialist, for her asthma. So Jim set up a master plan.

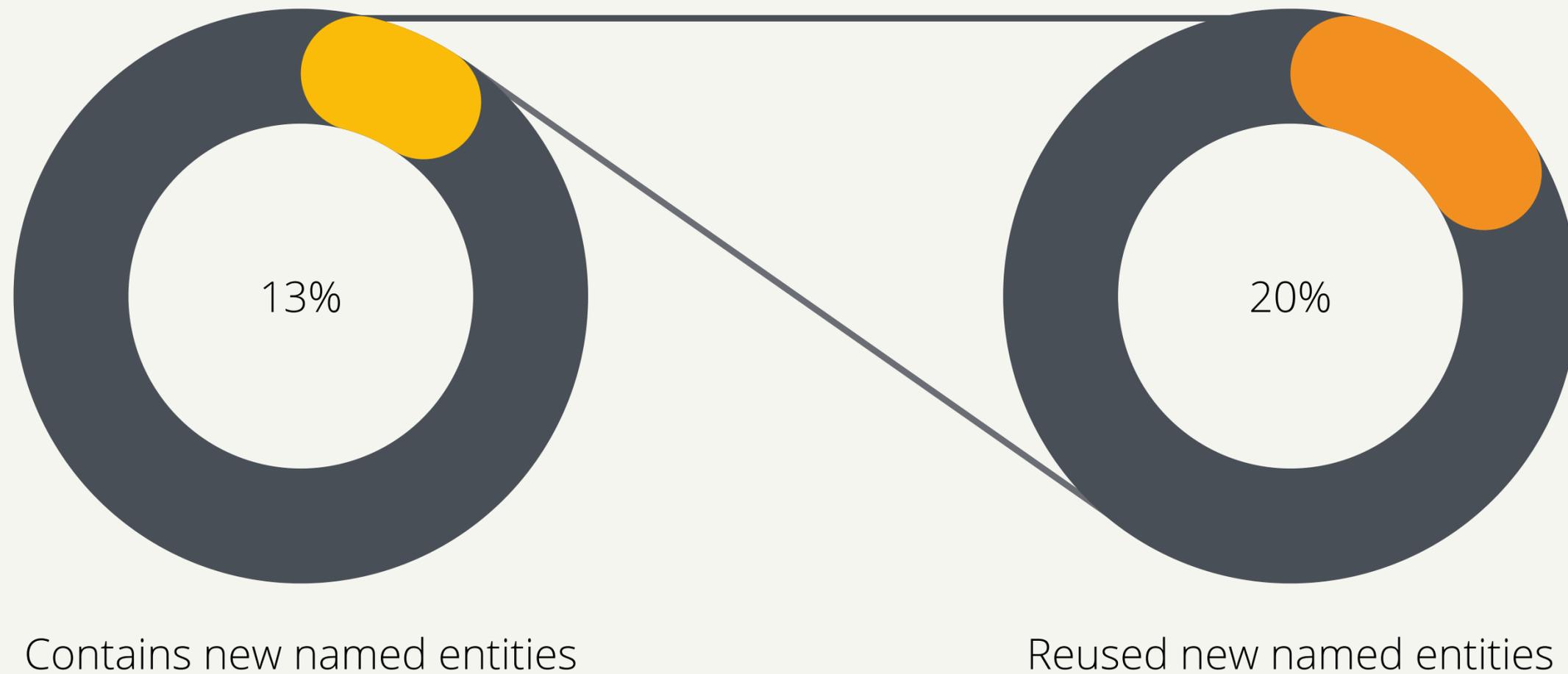
He applied for a job at the medical center, knowing that **Dr. John** had a vacation coming up. Jim got a job as an Asthma specialist, and he made sure he shadowed **Dr. John** with all of his patients so he would take over when he was on vacation. Karen happened to have an appointment the week **Dr. John** was going to be on vacation, so that was when Jim was going to try to swoop in.

It was the day of the appointment, and Jim was excited to set his plan in motion. [...]



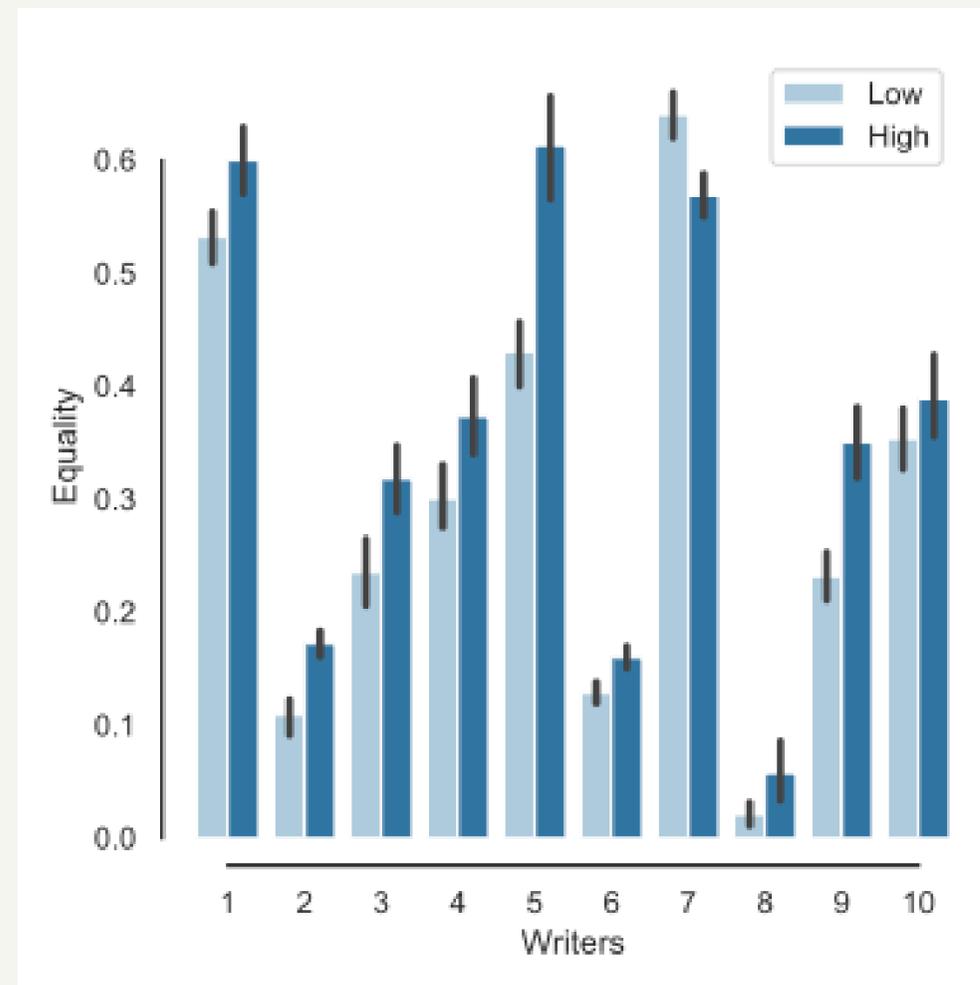
Ideation: Is GPT-3 capable of proposing **new ideas** for humans?

GPT-3 is **capable of providing new ideas** and influencing subsequent writing



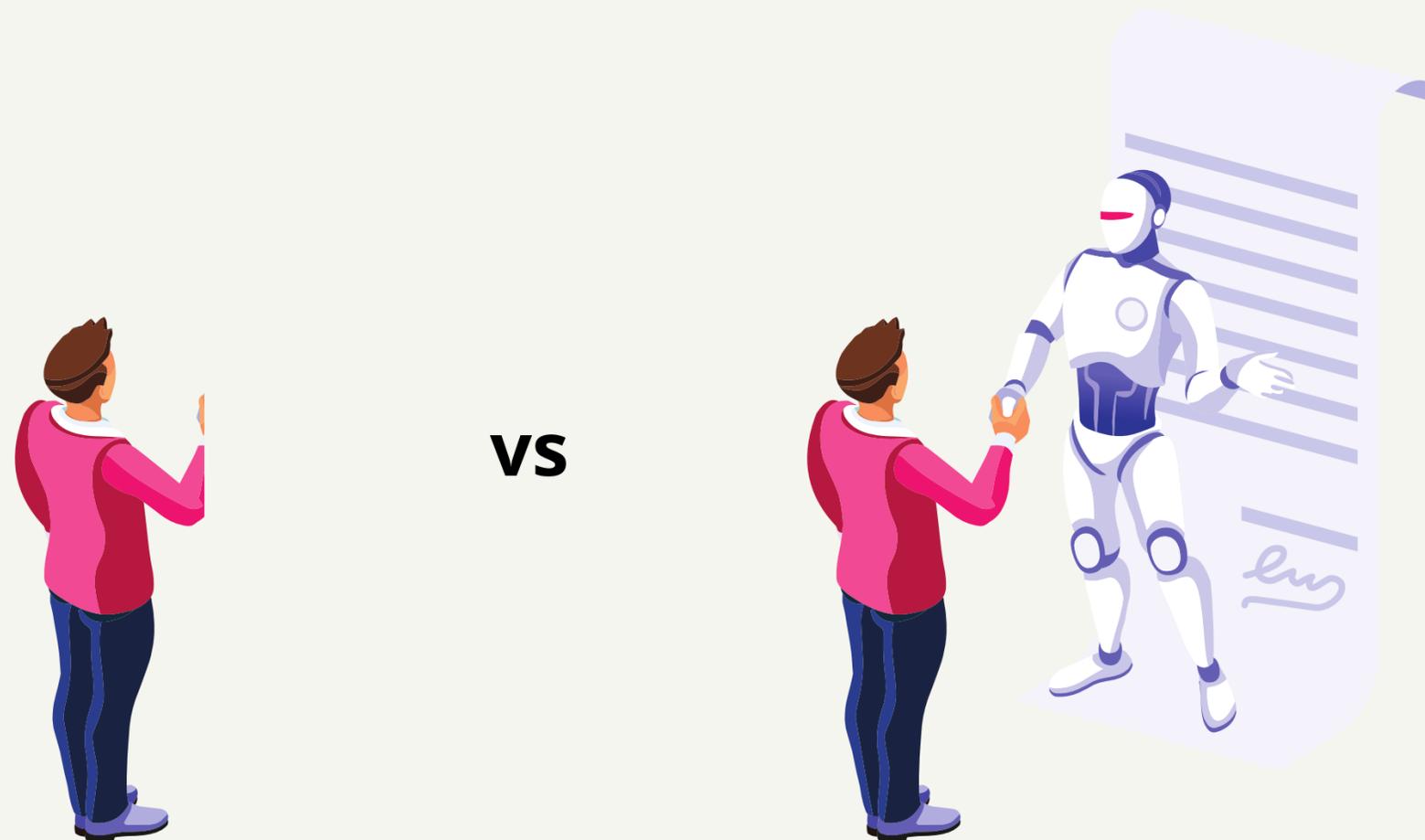
Collaboration: To what extent can GPT-3 collaborate with humans?

Individual writers have **different preference** when it comes to collaboration



Collaboration: How does the amount of collaboration impact humans' ownership and satisfaction?

The amount of collaboration with GPT-3 does not correlate with writers' **satisfaction** but negatively correlate with **ownership** over resultant text





Food for thought





Being creative requires a *jolt* to take us out of the well-carved paths we retrace each day.

This is where a machine might come in: perhaps it could give us that jolt, throw up a new suggestion, stop us from simply repeating the same algorithm each day.

Machines might ultimately help us, as humans, behave less like machines.

- Marcus Du Sautoy, *The Creativity Code: Art and Innovation in the Age of AI*

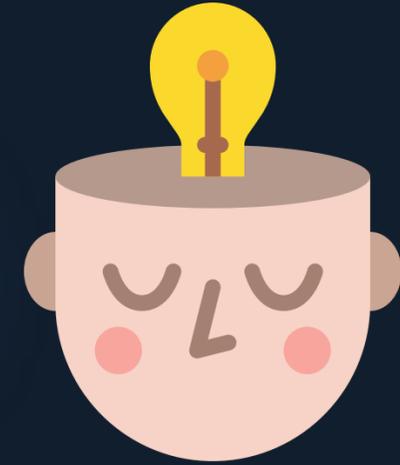


For those who choose to use artificial intelligence, it will alter the task of writing.

“The writer’s job becomes as an *editor* almost,” Gupta said. “Your role starts to become deciding what’s good and executing on your taste, not as much the low-level work of pumping out word by word by word. You’re still editing lines and copy and making those words beautiful, but, as you move up in that chain, and you’re executing your taste, you have the potential to do a lot more.”

- Stephen Marche, *The Computers Are Getting Better at Writing* (The New Yorker)

DISCUSSION



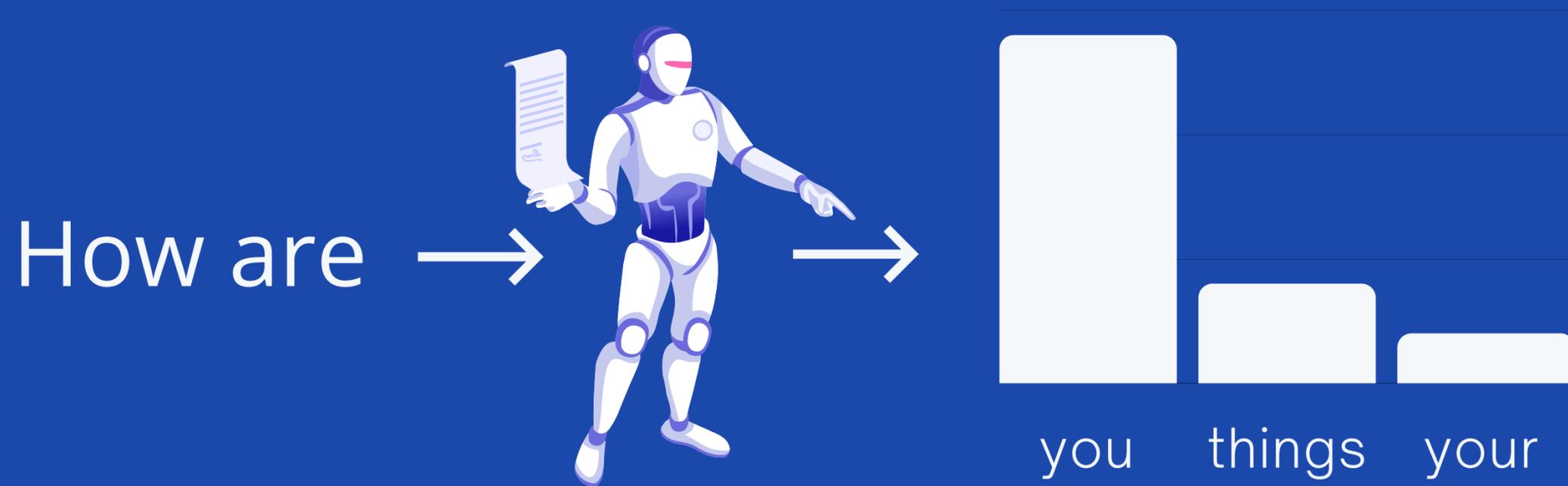
- Do you agree or disagree with this quote? Why?

“**The writer’s job becomes as an *editor* almost.** Your role starts to become deciding what’s good and executing on your taste, not as much the low-level work of pumping out word by word by word.”

- What do you think writing will be like in the future?
- What would be the most passive and proactive form of writing?

Summary

- We use **language models** everyday!
- Language models can offer unprecedented generative capabilities, which offers exciting opportunities for user experience design and innovation



Given a context, predict the next word

Summary

- You can design and **build writing assistants** with language models in many different ways (e.g. **keyword-based** autocomplete for **mobile environments**).
- Language models can be used as **contextual** thesaurus systems
- One way to **evaluate writing assistants** is building benchmarks (e.g. Swords)

Q. What are appropriate synonyms to replace the word "amazing" in the sentence "I read an amazing paper today"?

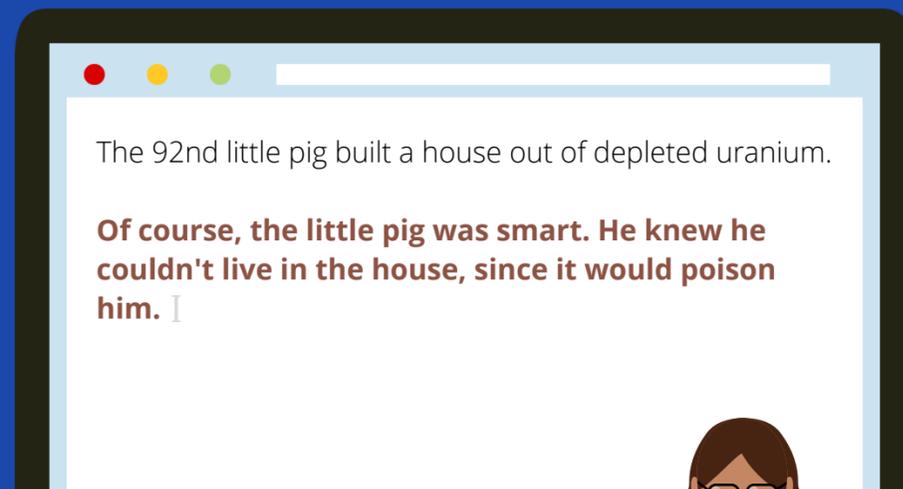
A. wonderful, astonishing, outstanding, stunning, astounding, fantastic, marvelous, breathtaking, astounding, astonishing, incredible, remarkable, incredible, wonderful, stunning, amazing, fabulous, sensational, surprising, staggering



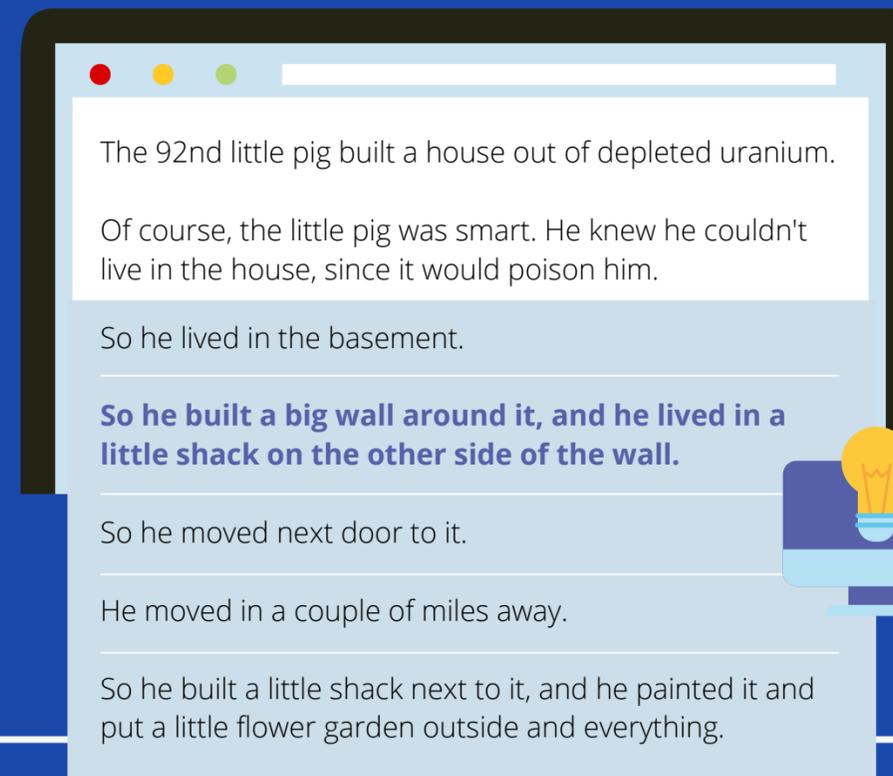
Summary

- We can reason about **interaction** between **humans** and **writing assistants** by recording all interactions in the text editor and analyzing logs
- Language models can provide **new ideas** to humans

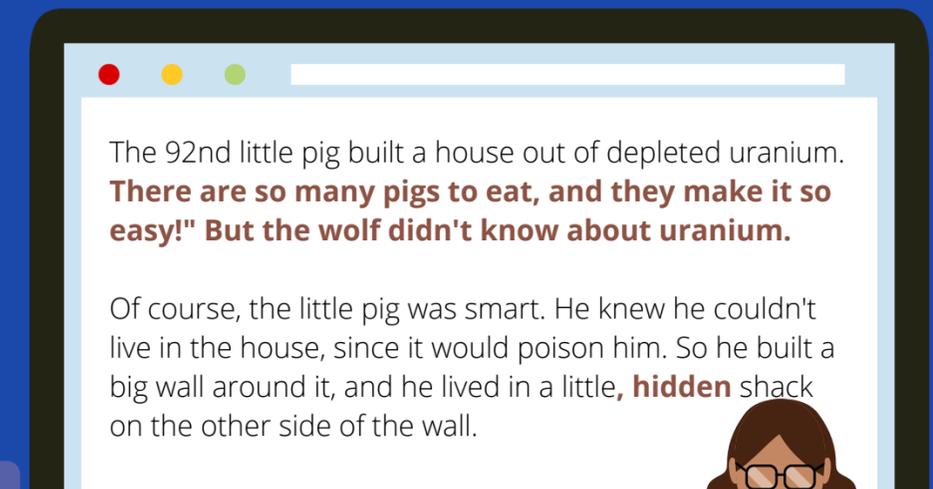
1. Write



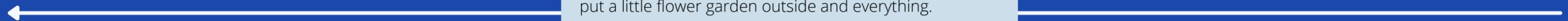
2. Get suggestions (optional)



3. Edit (optional)



Repeat





Thank you!

Mina Lee
minalee@cs.stanford.edu

