



LLMs: APPLICATIONS AND ETHICAL CONSIDERATIONS

Sandeep Soni

11/01/2023

PS2

- In section 2, take the top 20 genre features!
- Use pd.merge instead of df.merge to join two dataframes



- Kent Chang
- PhD student at UC Berkeley
- Researches social interaction in cultural texts
- Guest lecturer on Nov. 8

STORY SO FAR

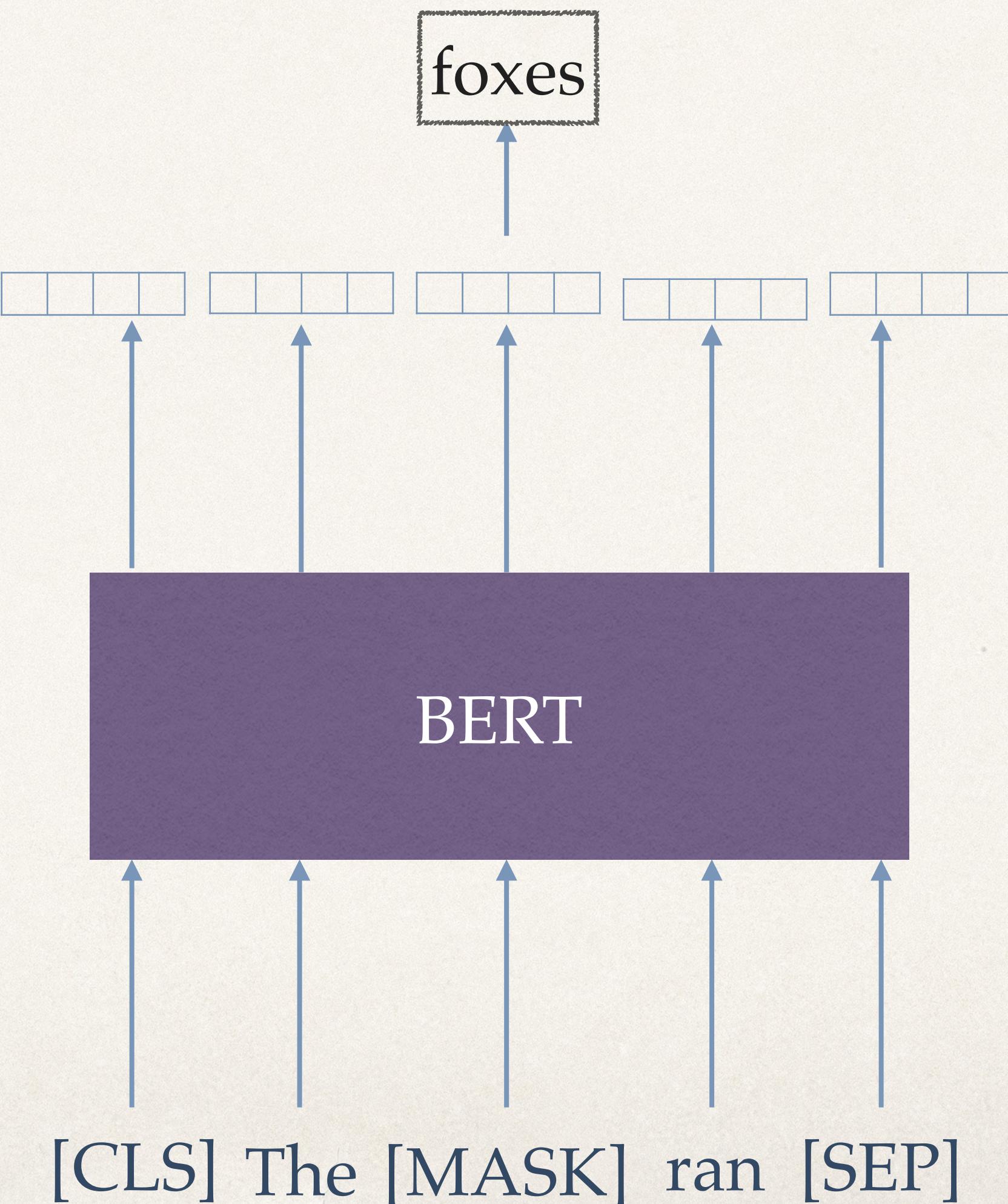
- **Language modeling**: Predict next word based on some previous context
i.e $P(x_i | x_1, x_2, \dots, x_{i-1})$
- **N-gram language models**: Predict next word based on only the n-1 previous words
 - e.g., trigram (n=3) $\rightarrow P(x_i | x_{i-2}, x_{i-1})$
- **RNN LM**: Predict the next word by calculating the similarity between word embeddings and context embeddings
 - $P(x_i | x_1, x_2, \dots, x_{i-1}) = \text{softmax}(\beta_{x_i} \cdot h_{i-1})$, where $h_{i-1} = \text{RNN}(x_{i-1}, h_{i-2})$

MASKED LANGUAGE MODELING

The _____ ran

Fill in the blank by
using the surrounding
context

$$P(w_t | \neg w_t)$$



BERT is a
language
model
trained to
predict the
missing
word

VARIATIONS OF LANGUAGE MODELING

Masked LM:

$$P(x) = \prod_{i=1}^n P(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

Encoder-Decoder LM:

$$P(y) = \prod_{i=1}^n P(y_i | y_1, \dots, y_{i-1}, x)$$

Causal LM:

$$P(x) = \prod_{i=1}^n P(x_i | x_1, x_2, \dots, x_{i-1})$$

BERT

- Masked LM

$$P(x) = \prod_{i=1}^n P(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

- The _____ ran → BERT → foxes

- Transformer based model

- trained on 512 sized contexts
- 12 or 24 transformer blocks
- 768 or 1024 sized representations
- pretrained on billions of tokens from wikipedia and books

T5

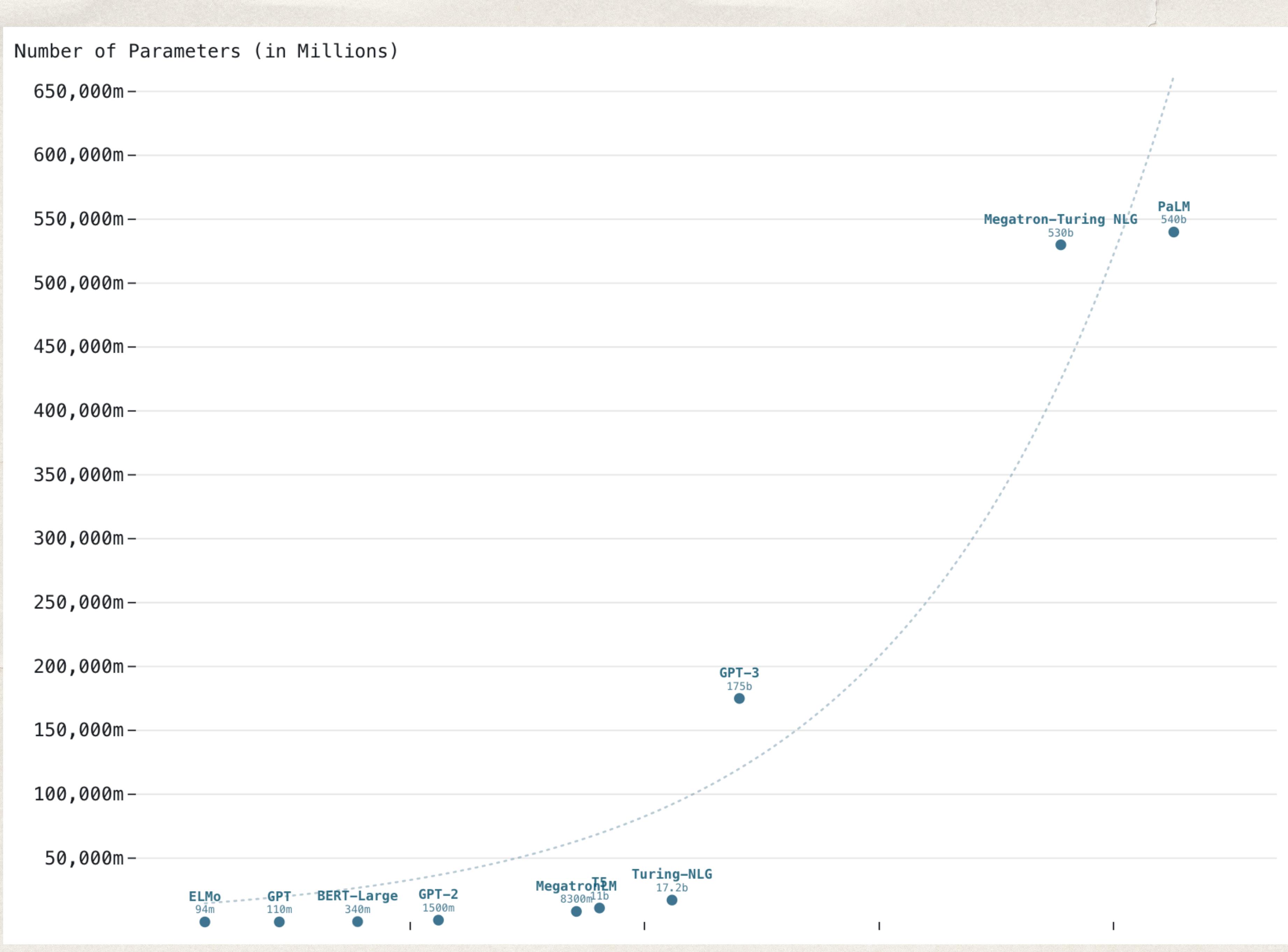
- Encoder-Decoder LM:
$$P(y) = \prod_{i=1}^n P(y_i | y_1, \dots, y_{i-1}, x)$$
- The ~~foxes ran in the forest after seeing the humans~~



- Pretrained on 750GB of web text

GPT

- Causal LM:
$$P(x) = \prod_{i=1}^n P(x_i | x_1, x_2, \dots, x_{i-1})$$
- Also called as left-to-right or autoregressive language modeling
- Fill in the blank but always predict the next word in the sequence (e.g., The foxes ____)
- GPT-3 has 175B parameters and trained on 570GB of web text, books, wikipedia



Modern LMs learn from massive amounts of text data and have rich representation

“What can we do with large language models?”

GENERATION

- Language models are generative models and hence they can be used to generate conversational text



I'm going to cook for my date who claims to be a picky eater. Can you recommend me a dish that's easy to cook?



Certainly! Cooking for a picky eater can be a bit of a challenge, but there are some dishes that tend to be universally well-received and are relatively easy to prepare. One option is to make a classic pasta dish, such as Spaghetti with Tomato Sauce and Meatballs. It's a familiar and comforting meal that most people enjoy. Here's a simple recipe:



Ingredients:

- 8 oz (about 225g) spaghetti
- 1 cup of your favorite tomato sauce (store-bought or homemade)
- 4-6 pre-cooked meatballs (you can buy frozen ones or make your own)
- Grated Parmesan cheese (optional, for topping)
- Fresh basil leaves (optional, for garnish)

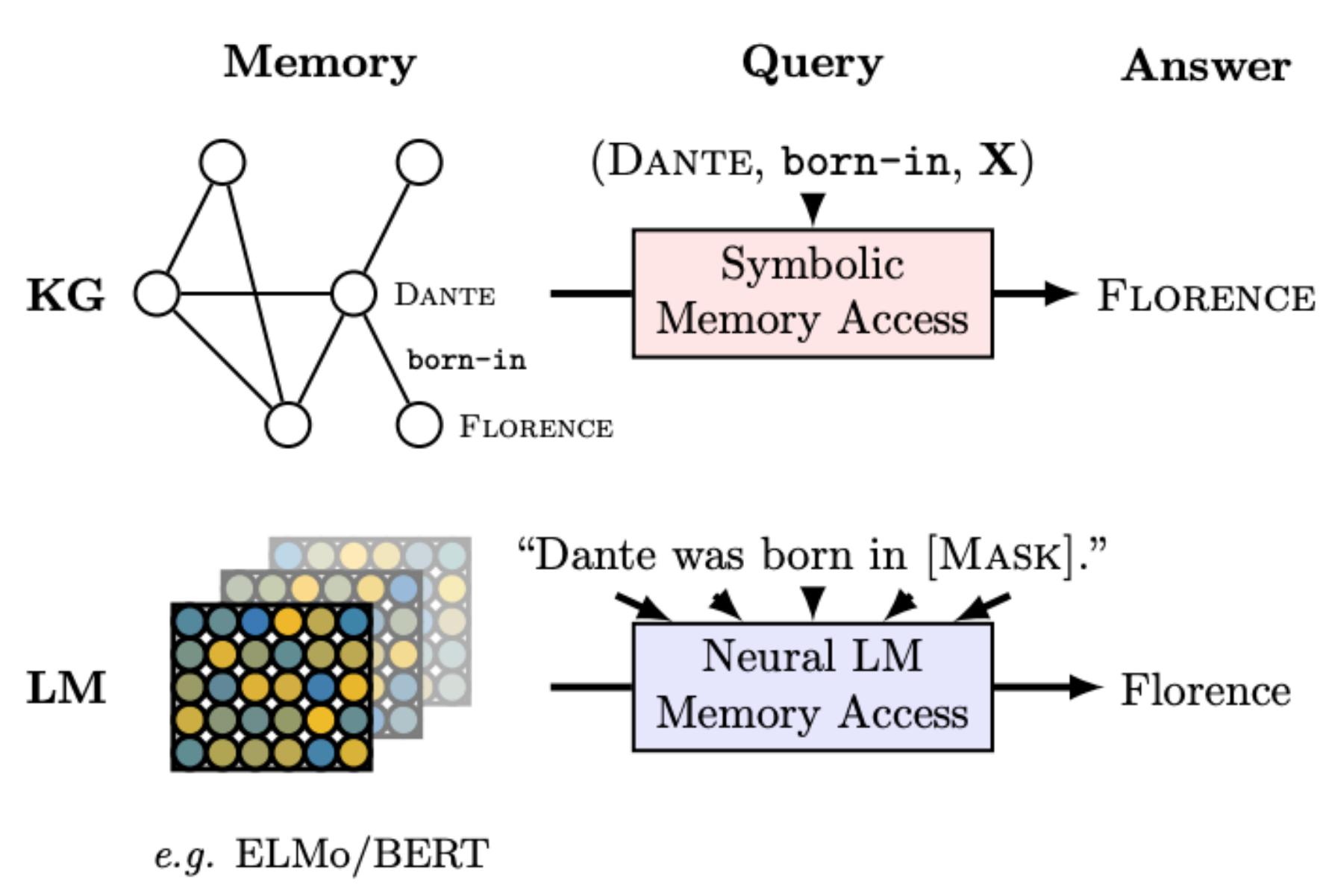
Instructions:

1. Cook the spaghetti according to the package instructions until al dente. Drain and set aside.
2. While the spaghetti is cooking, heat your tomato sauce in a saucepan over low heat. If your date is particularly picky, you can use a smooth tomato sauce without visible chunks of vegetables or herbs.
3. Heat the pre-cooked meatballs according to the package instructions. You can heat them in the tomato sauce or in the oven.
4. Once everything is ready, plate the cooked spaghetti, top it with the warm tomato sauce, and arrange the meatballs on top.
5. If your date enjoys cheese, you can sprinkle some grated Parmesan over the top. You can also add a few fresh basil leaves for a touch of freshness.

This simple Spaghetti with Tomato Sauce and Meatballs is a crowd-pleaser and can be customized to suit your date's preferences. If your date has specific dietary restrictions or dislikes certain ingredients, be sure to ask them in advance to make any necessary adjustments. Enjoy your meal, and I hope your date goes well!

Regenerate

LANGUAGE MODELS ARE KNOWLEDGE BASES



Query	Answer	Generation
Francesco Bartolomeo Conti was born in ____.	Florence	Rome [-1.8], Florence [-1.8], Naples [-1.9], Milan [-2.4], Bologna [-2.5]
Adolphe Adam died in ____.	Paris	Paris [-0.5], London [-3.5], Vienna [-3.6], Berlin [-3.8], Brussels [-4.0]
English bulldog is a subclass of ____.	dog	dogs [-0.3], breeds [-2.2], dog [-2.4], cattle [-4.3], sheep [-4.5]
The official language of Mauritius is ____.	English	English [-0.6], French [-0.9], Arabic [-6.2], Tamil [-6.7], Malayalam [-7.0]
Patrick Oboya plays in ____ position.	midfielder	centre [-2.0], center [-2.2], midfielder [-2.4], forward [-2.4], midfield [-2.7]
Hamburg Airport is named after ____.	Hamburg	Hess [-7.0], Hermann [-7.1], Schmidt [-7.1], Hamburg [-7.5], Ludwig [-7.5]
The original language of Mon Oncle Benjamin is ____.	French	French [-0.2], Breton [-3.3], English [-3.8], Dutch [-4.2], German [-4.9]
Dani Alves plays with ____.	Barcelona	Santos [-2.4], Porto [-2.5], Sporting [-3.1], Brazil [-3.3], Portugal [-3.7]
Paul Toungui is a ____ by profession.	politician	lawyer [-1.1], journalist [-2.4], teacher [-2.7], doctor [-3.0], physician [-3.7]
Sodium sulfide consists of ____.	sodium	water [-1.2], sulfur [-1.7], sodium [-2.5], zinc [-2.8], salt [-2.9]
Gordon Scholes is a member of the ____ political party.	Labor	Labour [-1.3], Conservative [-1.6], Green [-2.4], Liberal [-2.9], Labor [-2.9]
Kenya maintains diplomatic relations with ____.	Uganda	India [-3.0], Uganda [-3.2], Tanzania [-3.5], China [-3.6], Pakistan [-3.6]
iPod Touch is produced by ____.	Apple	Apple [-1.6], Nokia [-1.7], Sony [-2.0], Samsung [-2.6], Intel [-3.1]
Bailey Peninsula is located in ____.	Antarctica	Antarctica [-1.4], Bermuda [-2.2], Newfoundland [-2.5], Alaska [-2.7], Canada [-3.1]
JDK is developed by ____.	Oracle	IBM [-2.0], Intel [-2.3], Microsoft [-2.5], HP [-3.4], Nokia [-3.5]
Carl III used to communicate in ____.	Swedish	German [-1.6], Latin [-1.9], French [-2.4], English [-3.0], Spanish [-3.0]
Sunshine Coast, British Columbia is located in ____.	Canada	Canada [-1.2], Alberta [-2.8], Yukon [-2.9], Labrador [-3.4], Victoria [-3.4]
Pope Clement VII has the position of ____.	pope	cardinal [-2.4], Pope [-2.5], pope [-2.6], President [-3.1], Chancellor [-3.2]
Joe Cocker is represented by music label ____.	Capitol	EMI [-2.6], BMG [-2.6], Universal [-2.8], Capitol [-3.2], Columbia [-3.3]
London Jazz Festival is located in ____.	London	London [-0.3], Greenwich [-3.2], Chelsea [-4.0], Camden [-4.6], Stratford [-4.8]
Border TV is owned by ____.	ITV	Sky [-3.1], ITV [-3.3], Global [-3.4], Frontier [-4.1], Disney [-4.3]
The native language of Mammootty is ____.	Malayalam	Malayalam [-0.2], Tamil [-2.1], Telugu [-4.8], English [-5.2], Hindi [-5.6]
The Sharon Cuneta Show was created in ____.	Philippines	Manila [-3.2], Philippines [-3.6], February [-3.7], December [-3.8], Argentina [-4.0]
You are likely to find a overflow in a ____.	drain	sewer [-3.1], canal [-3.2], toilet [-3.3], stream [-3.6], drain [-3.6]
Ravens can ____.	fly	fly [-1.5], fight [-1.8], kill [-2.2], die [-3.2], hunt [-3.4]
Joke would make you want to ____.	laugh	cry [-1.7], die [-1.7], laugh [-2.0], vomit [-2.6], scream [-2.6]
Sometimes virus causes ____.	infection	disease [-1.2], cancer [-2.0], infection [-2.6], plague [-3.3], fever [-3.4]
Birds have ____.	feathers	wings [-1.8], nests [-3.1], feathers [-3.2], died [-3.7], eggs [-3.9]
Typing requires ____.	speed	patience [-3.5], precision [-3.6], registration [-3.8], accuracy [-4.0], speed [-4.1]
Time is ____.	finite	short [-1.7], passing [-1.8], precious [-2.9], irrelevant [-3.2], gone [-4.0]
You would celebrate because you are ____.	alive	happy [-2.4], human [-3.3], alive [-3.3], young [-3.6], free [-3.9]
Skills can be ____.	taught	acquired [-2.5], useful [-2.5], learned [-2.8], combined [-3.9], varied [-3.9]
A pond is for ____.	fish	swimming [-1.3], fishing [-1.4], bathing [-2.0], fish [-2.8], recreation [-3.1]

EVERYTHING IS LANGUAGE MODELING

- If you can construct a suitable prompt, many tasks can be framed as language modeling
 - The director of Avatar is _____
 - The definition of “chromatid” is _____
 - Is “sit down” rude than “please, sit down”? _____

PRETRAINED AND FINETUNED

- Though a language model is trained on predicting words in sequences, a pretrained language model can be adapted for a specific task – a process called finetuning
- Because of the size of some LMs, finetuning can be costly

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

- 1 Translate English to French: ← task description
- 2 cheese => ← prompt

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

- 1 Translate English to French: ← task description
- 2 sea otter => loutre de mer ← example
- 3 cheese => ← prompt

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

- 1 Translate English to French: ← task description
- 2 sea otter => loutre de mer ← examples
- 3 peppermint => menthe poivrée
- 4 plush girafe => girafe peluche
- 5 cheese => ← prompt

Language Models are Few-Shot Learners:
<https://arxiv.org/pdf/2005.14165.pdf>

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. X

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓

Using GPT-4 to measure the passage of time in fiction

Large language models are valuable research assistants,
especially when they refuse to follow instructions.

<https://tedunderwood.com/2023/03/19/using-gpt-4-to-measure-the-passage-of-time-in-fiction/>

What is the amount of narrative time that has passed in a short excerpt of fictional text?

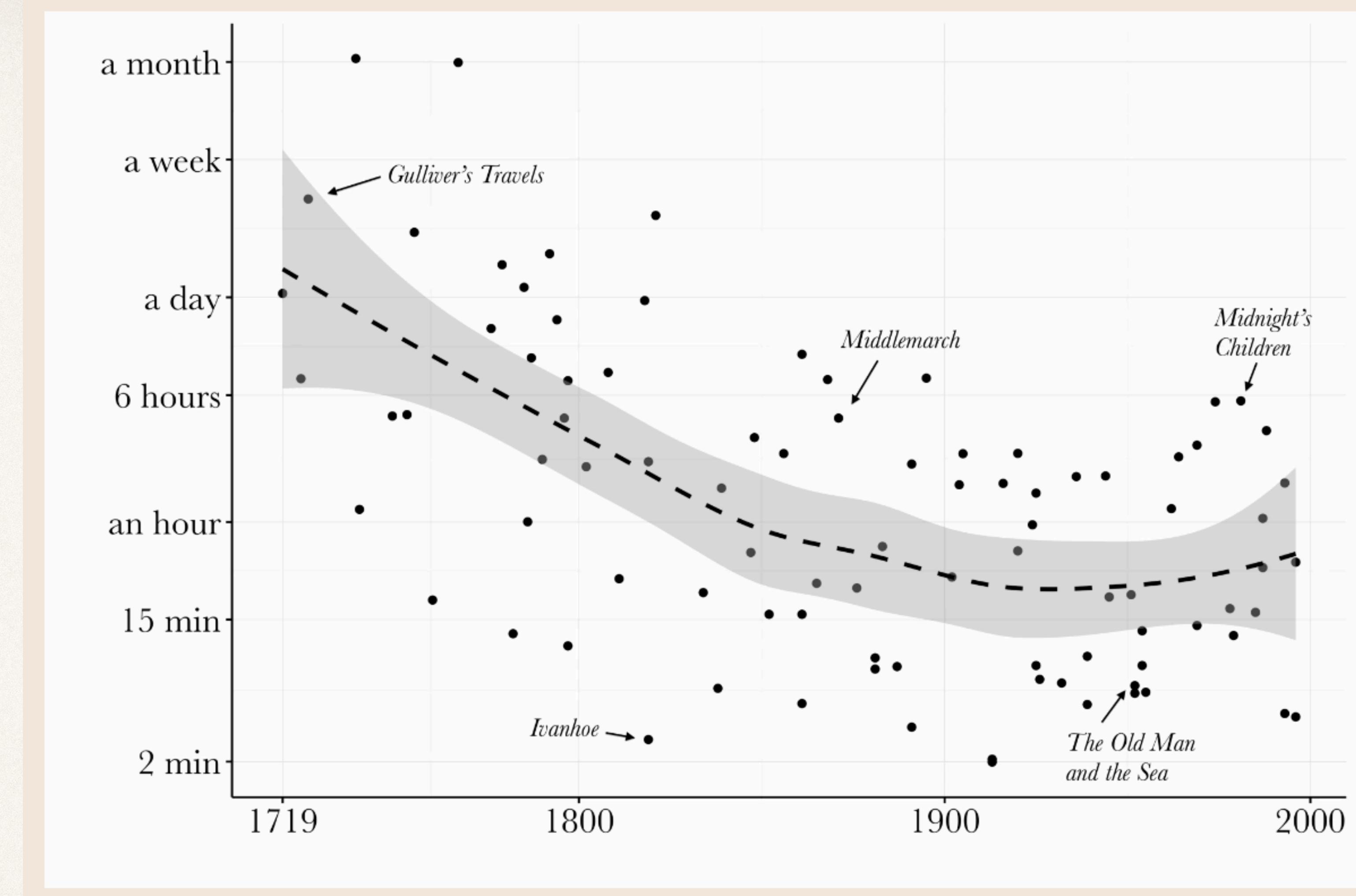


Image taken from Underwood's article "Why Literary Time is Measured in Minutes"

Read the following passage of fiction. Then do five things.

- 1: Briefly summarize the passage.
 - 2: Reason step by step to decide how much time is described in the passage. If the passage doesn't include any explicit reference to time, you can guess how much time the events described would have taken. Even description can imply the passage of time by describing the earlier history of people or buildings. But characters' references to the past or future in spoken dialogue should not count as time that passed in the scene. Report the time using units of years, weeks, days, hours, or minutes. Do not say zero or N/A.
 - 3: If you described a range of possible times in step 2 take the midpoint of the range. Then multiply to convert the units into minutes.
 - 4: Report only the number of minutes elapsed, which should match the number in step 3. Do not reply N/A.
 - 5: Given the amount of speculation required in step 2, describe your certainty about the estimate--either high, moderate, or low.
- The passage follows: <omitting this to save space>

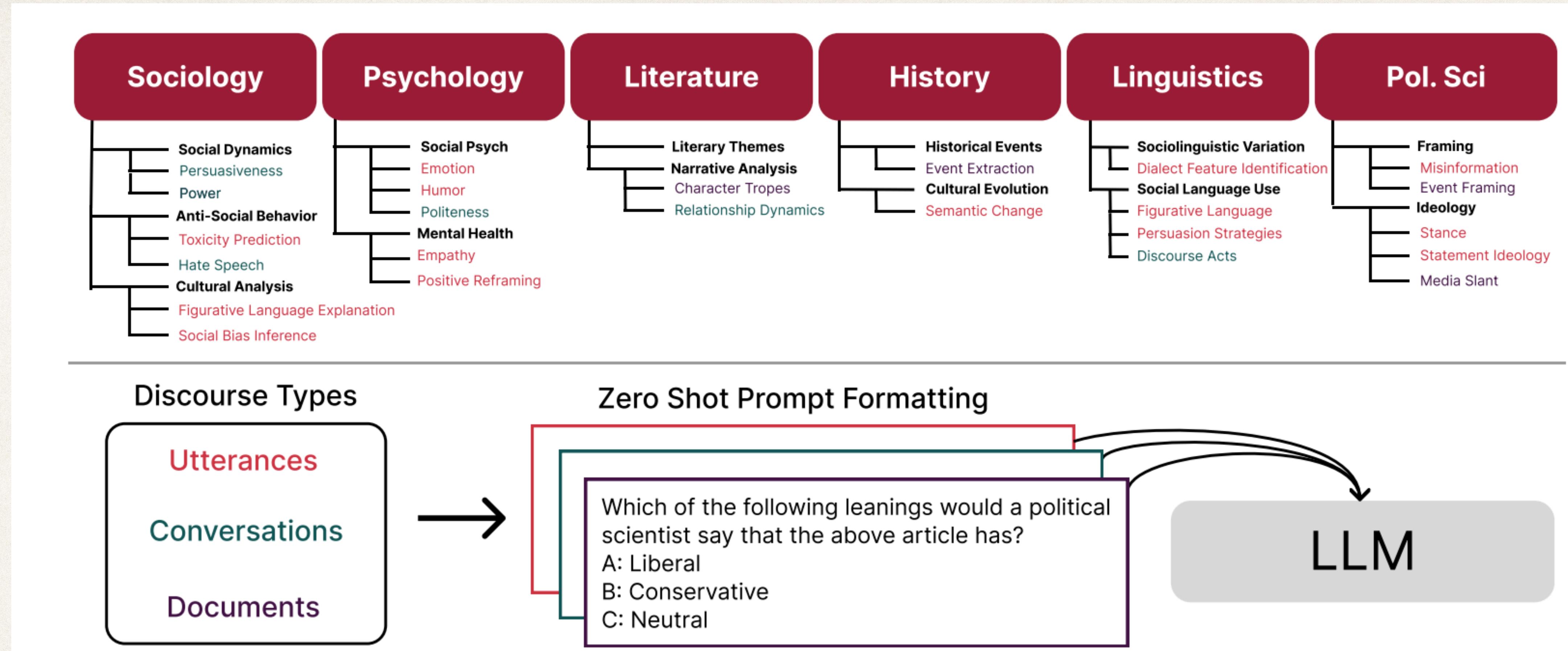
- 1: A detective wakes up 'the next morning,' looks out a window for an undefined time, drinks (and presumably needs to make) two cups of coffee, then shaves and showers and gets dressed before stepping out his front door and seeing a car.
- 2: Making coffee, showering, and getting dressed take at least an hour. There's some ambiguity about whether to count the implicit reference to yesterday (since this is 'the next morning') as time elapsed in the passage, but let's say no, since yesterday is not actually described. So, an hour to 90 minutes.
- 3: 1.25 hours have elapsed. Multiplying by 60 minutes an hour that's 75 minutes.
- 4: 75 minutes.
- 5: Low confidence, because of ambiguity about a reference to the previous day.

Response from GPT-4

Table 1: Comparison of modeling methods

Modeling method	Features	Communication between readers	Pearson's r between readers
Linear regression	words	no	.35
Ridge regression	words	yes	.49
ChatGPT	full text	yes	.59
GPT-4	full text	yes	.68
Human reading	full text	yes	.74

LLMs are promising “research assistants” on measurement tasks



Ziems, Caleb, et al. "Can Large Language Models Transform Computational Social Science?."

Large language models can be effective partners to humans on multiple tasks

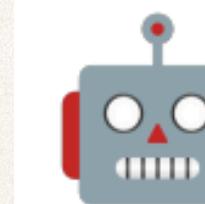
MULTIMODALITY

Language models can be integrated with other modalities (e.g., vision)

c)



Who are the people on the right, on the left, and in the middle?



The people in the image are:

- * Barack Obama, the former President of the United States, is on the left side of the image.
- * Joe Biden, the current President of the United States, is in the middle of the image.
- * Donald Trump, the former President of the United States, is on the right side of the image.

“Are there any risks?”

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Emily M. Bender*

ebender@uw.edu

University of Washington

Seattle, WA, USA

Timnit Gebru*

timnit@blackinai.org

Black in AI

Palo Alto, CA, USA

Angelina McMillan-Major

aymm@uw.edu

University of Washington

Seattle, WA, USA

Shmargaret Shmitchell

shmargaret.shmitchell@gmail.com

The Aether

We should weigh the financial cost and environmental risks associated with LLMs

We should understand more rigorously the data on which LLMs are trained and carefully document these datasets

HALLUCINATIONS

- Large Language models are capable of producing fluent text but not necessarily factually accurate text
- LLMs can make up stuff and still sound confident
- This problem is called hallucination

Lawyer cites fake cases generated by ChatGPT in legal brief

The high-profile incident in a federal case highlights the need for lawyers to verify the legal insights generated by AI-powered tools.

Published May 30, 2023



Lyle Moran
Reporter

[in](#) [f](#) [t](#) [p](#) [m](#)



ENERGY CONSUMPTION

- Training large language models isn't just financially expensive, it is hurting the environment
- Potential solutions is to reduce the carbon footprint by making the models more efficient
- At the very least though we should highlight these environmental harms

Consumption	CO ₂ e (lbs)
Air travel, 1 person, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL) w/ tuning & experiments	39 78,468
Transformer (big) w/ neural arch. search	192 626,155

Table 1: Estimated CO₂ emissions from training common NLP models, compared to familiar consumption.¹

Strubell et. al. 2019

TRANSPARENCY

- Many things about some LLMs are unknown (e.g., training data, training protocol, etc)
- How can we know that LLMs are generalizing and not just memorizing?
- These can have implications on the confidence we have in our measurements based on LLMs

Wow. I sit down, fish the questions from my backpack, and go through them, inwardly cursing [MASK] for not providing me with a brief biography. I know nothing about this man I'm about to interview. He could be ninety or he could be thirty. → **Kate** (James, *Fifty Shades of Grey*).

Some days later, when the land had been moistened by two or three heavy rains, [MASK] and his family went to the farm with baskets of seed-yams, their hoes and machetes, and the planting began. → **Okonkwo** (Achebe, *Things Fall Apart*).

Figure 1: Name cloze examples. GPT-4 answers both of these correctly.

Figure taken from Chang et. al. 2023

LLMs can be highly accurate in measurement tasks on data already seen in training but may not generalize for unseen data

If LLMs are memorizing data from contemporary books, what does it tell you about copyright and fair use?

GPT-4	ChatGPT	BERT	Date	Author	Title
0.98	0.82	0.00	1865	Lewis Carroll	<i>Alice's Adventures in Wonderland</i>
0.76	0.43	0.00	1997	J.K. Rowling	<i>Harry Potter and the Sorcerer's Stone</i>
0.74	0.29	0.00	1850	Nathaniel Hawthorne	<i>The Scarlet Letter</i>
0.72	0.11	0.00	1892	Arthur Conan Doyle	<i>The Adventures of Sherlock Holmes</i>
0.70	0.10	0.00	1815	Jane Austen	<i>Emma</i>
0.65	0.19	0.00	1823	Mary W. Shelley	<i>Frankenstein</i>
0.62	0.13	0.00	1813	Jane Austen	<i>Pride and Prejudice</i>
0.61	0.35	0.00	1884	Mark Twain	<i>Adventures of Huckleberry Finn</i>
0.61	0.30	0.00	1853	Herman Melville	<i>Bartleby, the Scrivener</i>
0.61	0.08	0.00	1897	Bram Stoker	<i>Dracula</i>
0.61	0.18	0.00	1838	Charles Dickens	<i>Oliver Twist</i>
0.59	0.13	0.00	1902	Arthur Conan Doyle	<i>The Hound of the Baskervilles</i>
0.59	0.22	0.00	1851	Herman Melville	<i>Moby Dick; Or, The Whale</i>
0.58	0.35	0.00	1876	Mark Twain	<i>The Adventures of Tom Sawyer</i>
0.57	0.30	0.00	1949	George Orwell	<i>1984</i>
0.54	0.10	0.00	1908	L. M. Montgomery	<i>Anne of Green Gables</i>
0.51	0.20	0.01	1954	J.R.R. Tolkien	<i>The Fellowship of the Ring</i>
0.49	0.16	0.13	2012	E.L. James	<i>Fifty Shades of Grey</i>
0.49	0.24	0.01	1911	Frances H. Burnett	<i>The Secret Garden</i>
0.49	0.12	0.00	1883	Robert L. Stevenson	<i>Treasure Island</i>
0.49	0.16	0.00	1847	Charlotte Brontë	<i>Jane Eyre: An Autobiography</i>
0.49	0.22	0.00	1903	Jack London	<i>The Call of the Wild</i>

Table 1: Top 20 books by GPT-4 name cloze accuracy.

IN CLASS EXERCISE

- LLMs prompting