

# Examining the Reversal Curse

---

Mona Gandhi

CIS 7000

Project 1

02/26/24

# Motivation

For humans, **if A is B  $\rightarrow$  B is A.**

Example: Joe Biden is the President of USA  $\rightarrow$  The President of USA is Joe Biden.

The paper '*The Reversal Curse: LLMs trained on "A is B" fail to learn "B is A"*' brings to notice that (GPT) models fail on such tasks where it has never seen B is A.

# Replicating results from the paper

Prompts:

Ask for parent prompt: "Who is X's Y?"  $\rightarrow$  Z

Example: Who is Chris Hemsworth's father?  $\rightarrow$  Craig Hemsworth

Ask for child prompt: "Name a child of Z."  $\rightarrow$  X

Example: Name a child of Craig Hemsworth.  $\rightarrow$  Chris Hemsworth

Here,

X: celebrity

Y: parent type (mother/father)

Z: corresponding parent.

Method:

Few-shot prompting

Repeat the same prompt 10 times

# Replicating results from the paper

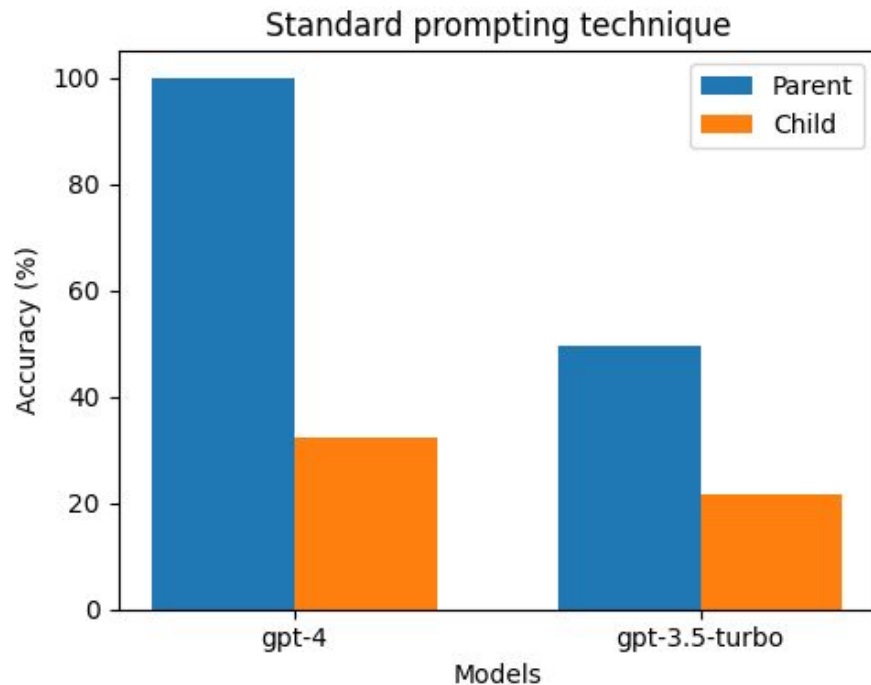
GPT 4: used to get ground truth parent names for celebrities

Number of parent child pairs: 1515

Testing sets:

Asking for child: GPT-4, GPT-3.5-turbo

Asking for parent: GPT-3.5-turbo



# Contradiction

A is B

**M You**  
Who is Tom Cruise's mother?

**ChatGPT**  
Tom Cruise's mother is Mary Lee Pfeiffer. She was born Mary Lee South and was a special education teacher. Tom Cruise, whose full name is Thomas Cruise Mapother IV, is a well-known American actor and producer.

B is not A

**M You**  
Who is Mary Lee Pfeiffer's son?

**ChatGPT**  
As of my last knowledge update in January 2022, I don't have information on Mary Lee Pfeiffer or her son. It's possible that she is a private individual or not widely known in public domains. If there have been developments or new information after my last update, I recommend checking the latest sources or conducting an online search for the most recent information. Keep in mind that details about private individuals may not be readily available or may be intentionally kept private.

Even contradicts itself in the same answer ... !!

**M You**  
Is Tom Cruise Mary Lee Pfeiffer's son?

**ChatGPT**  
No, Tom Cruise is not Mary Lee Pfeiffer's son. Tom Cruise was born on July 3, 1962, in Syracuse, New York, to Mary Lee Pfeiffer and Thomas Cruise Mapother III. Mary Lee Pfeiffer is Tom Cruise's mother, but Tom Cruise is not the son of actress Michelle Pfeiffer, who is a different person. Michelle Pfeiffer was born on April 29, 1958, in Santa Ana, California.

# Verification Prompting

Additional Information for verification.  
Example: Yes. Craig Hemsworth is Chris Hemsworth's father. Chris Hemsworth is Craig Hemsworth's son.

Prompts:

Ask for parent prompt: "Is Z X's Y?" → Yes. ...

Example: Is Craig Hemsworth Chris Hemsworth's father? → Yes. ...

Ask for child prompt: "Is X Z's child?" → Yes. ...

Example: Is Chris Hemsworth Craig Hemsworth's child? → Yes. ...

Here,

X: celebrity

Y: parent type (mother/father)

Z: corresponding parent.

Method:

Few-shot prompting

Repeat the same prompt 5 times

# Verification Results

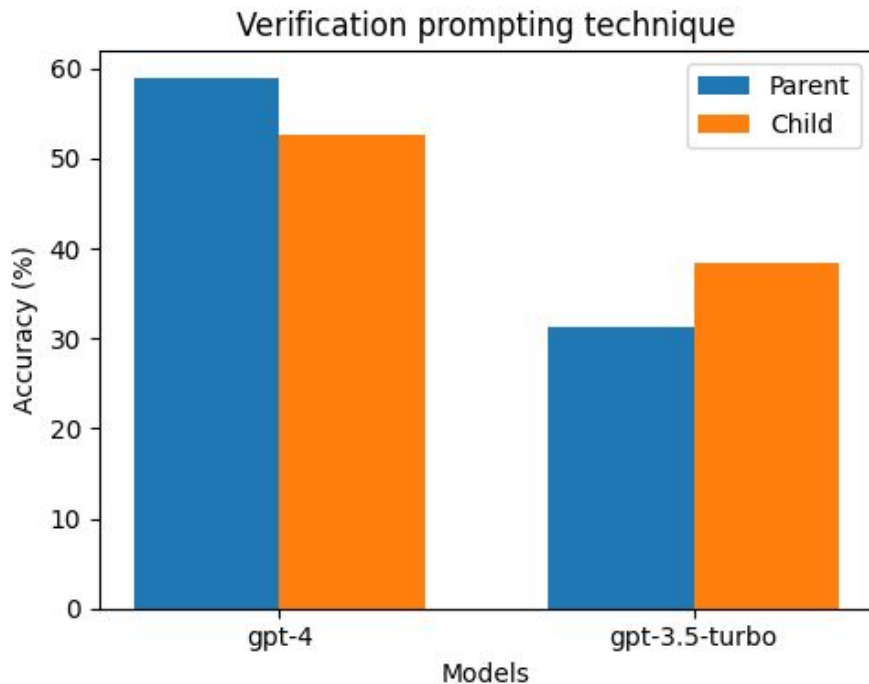
Expected answer of all prompts is starting with “Yes”.

Ground Truth based on results from GPT-4 (same as previous).

Testing sets:

Asking for child: GPT-4, GPT-3.5-turbo

Asking for parent: GPT-4, GPT-3.5-turbo



# Contradiction Experiment

Main question: How to evaluate?

1. A small human evaluation (~ 400 responses) ask a response from GPT having contradiction as 1 and 0 otherwise.
2. Use the created dataset to fine-tune BERT model for classification.
3. Evaluate all the responses using this model.

WHY BERT?

Learning rate:  $2e-5$

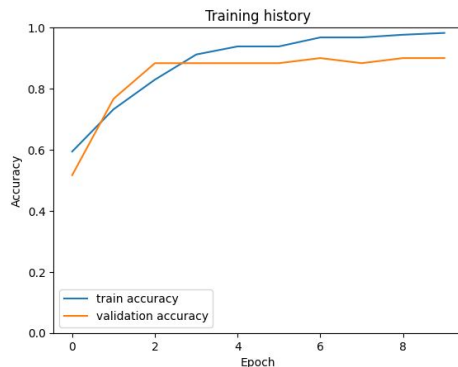
Epochs: 10

Loss: Cross Entropy Loss

Optimizer: AdamW

Training set size: 340

Val set size: 60 (balanced)



	Accuracy
Train	97.64%
Val	90.00%



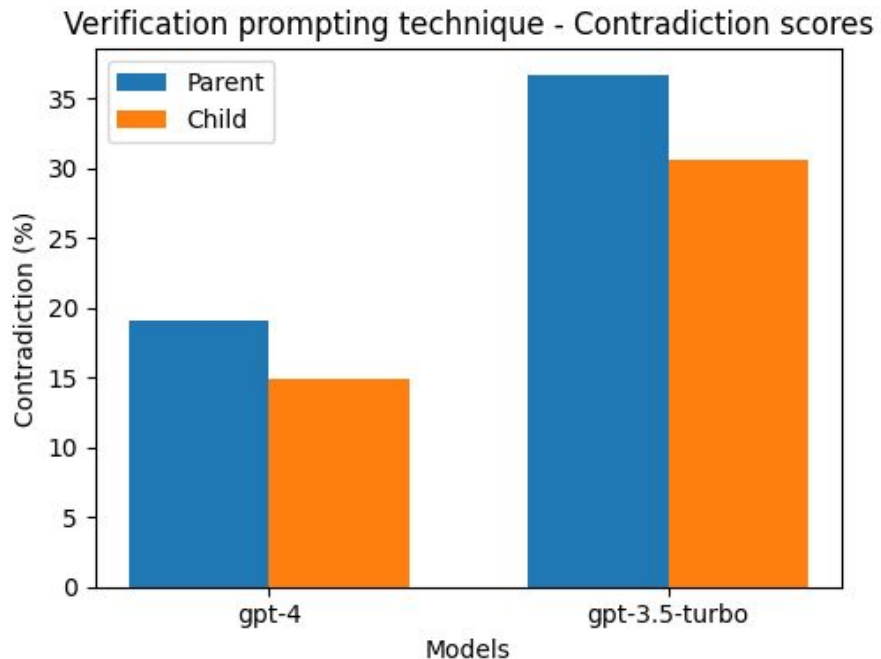
# Contradiction Results

Testing sets:

Asking for child: GPT-4, GPT-3.5-turbo

Asking for parent: GPT-3.5-turbo

Evaluation: BERT model



# Some Examples

Ground truth: Aaron Taylor-Johnson is Robert Johnson's father.

Query: Is Aaron Taylor-Johnson Robert Johnson's father?

Response: No. Robert Johnson is not related to Aaron Taylor-Johnson. Aaron Taylor-Johnson's father is Robert Johnson.

Ground truth: Dennis Page is Elliot Page's father.

Query: Is Dennis Page Elliot Page's father?

Response: No. Elliot Page's parents are Martha Phillipotts and Dennis Page. Dennis Page is not known to be related to Elliot Page.

# With google search! (*Perplexity AI*)

A is B

Who is Tom Cruise's mother?

## Sources

Tom Cruise - Wikipedia

wikipedia · 1

Tom Cruise's Mother, Mary Lee South, Dies

people · 2

Tom Cruise's mother has passed away at age 80

go · 3



View 2 more

## Answer

Tom Cruise's mother was Mary Lee South. She passed away at the age of 80. She was a supportive figure in Cruise's life, encouraging his acting career and providing unconditional support. Cruise was known for his unwavering devotion to his mother, and he acknowledged that his success was largely due to the values instilled in him by her, including a strong work ethic, determination, and patience <sup>1</sup> <sup>2</sup> <sup>3</sup> <sup>4</sup> <sup>5</sup>.

B is A

Who is Mary Lee Pfeiffer's son?

## Sources

Language models know Tom Cruise's mother, bu...

the-decoder · 1

Tom Cruise's Mother, Mary Lee South, Dies

people · 2

'They were very close': Tom Cruise was a...

meaww · 3



View 2 more

## Answer

Mary Lee Pfeiffer's son is Tom Cruise. Tom Cruise is a well-known Hollywood actor, and his mother, Mary Lee Pfeiffer, encouraged him to pursue his acting career. She passed away in 2017 at the age of 80.

Is Tom Cruise Mary Lee Pfeiffer's son?

## Sources

Language models know Tom Cruise's mother, bu...

the-decoder · 1

Tom Cruise's Mother, Mary Lee South, Dies

people · 2

Tom Cruise - Wikipedia

wikipedia · 3



View 2 more

## Answer

Tom Cruise is not Mary Lee Pfeiffer's son. Mary Lee Pfeiffer was the mother of Tom Cruise, born as Thomas Cruise Mapother IV. She was married to Tom Cruise's father, Thomas Cruise Mapother III, and they had four children together: Tom, Lee Anne, Marian, and Cass. Mary Lee Pfeiffer passed away in 2017.

Still contradicts itself in the same answer ... !!

# Manually tested results



Q: Is Dennis Page Elliot Page's father?

A: No, Dennis Page is not Elliot Page's father. Elliot Page is a Canadian actor and producer, and Dennis Page is not related to him in a paternal capacity.



Q: Is Aaron Taylor-Johnson Robert Johnson's father?

A: No, Aaron Taylor-Johnson is not Robert Johnson's father. Aaron Taylor-Johnson is an English actor known ...



Q: Is Michael Gadot Gal Gadot's father?

A: No, Michael Gadot is not Gal Gadot's father. Gal Gadot's father is Michael Gadot



Q: Is Martha Philpotts Elliot Page's mother?

A: No, Martha Philpotts is not Elliot Page's mother. Martha Philpotts is Elliot Page's mother



Q: Is Aridio Saldaña Zoe Saldana's father?

A: Yes. Aridio Saldaña is Zoe Saldana's father. Zoe Saldana is Aridio Saldaña's child.

# Manually tested results

 Q: Is Wes Anderson Melver Leonard Anderson's child?

A: No, Wes Anderson is not Melver Leonard Anderson's child. Wes Anderson is an American filmmaker ...

 Q: Is Aaron Paul Robert Sturtevant's child?

A: No. Aaron Paul is not Robert Sturtevant's child. Aaron Paul's father is Robert Sturtevant. Aaron Paul is Robert Sturtevant's child.

 Q: Is Ezra Miller Robert S. Miller's child?

A: No, Ezra Miller is not Robert S. Miller's child. There is no known familial relationship between Ezra Miller and Robert S. Miller based on available search results.

 Q: Is Tom Cruise Mary Lee Pfeiffer's child?

A: No, Tom Cruise is not Mary Lee Pfeiffer's child. Mary Lee Pfeiffer is Tom Cruise's mother

 Q: Is Olga Kurylenko Konstantin Kurylenko's child?

A: No, Olga Kurylenko is not Konstantin Kurylenko's child. The search results provided do not offer relevant information to confirm their relationship.

Figures for Report

1. Standard Prompt (A → B)



Who is Tom Cruise's mother?



Tom Cruise's mother is Mary Lee Pfeiffer [...]



2. Standard Prompt (B → A)



Who is Mary Lee Pfeiffer's son?



As of [...] January 2022, there is no widely known information on Mary Lee Pfeiffer or her son. It may be [...]



3. Verification Prompt



Is Tom Cruise Mary Lee Pfeiffer's son?



No, Tom Cruise is not Mary Lee Pfeiffer's son. Tom Cruise was [...]. Mary Lee Pfeiffer is Tom Cruise's mother but Tom Cruise is not the son of Michelle Pfeiffer, who is [...]



4. Verification Prompt (Perplexity AI)



Is Tom Cruise Mary Lee Pfeiffer's son?



Tom Cruise is not Mary Lee Pfeiffer's son. Mary Lee Pfeiffer was the mother of Tom Cruise, born as Thomas Cruise IV. She was married to Tom Cruise's father, [...]



NOTE:



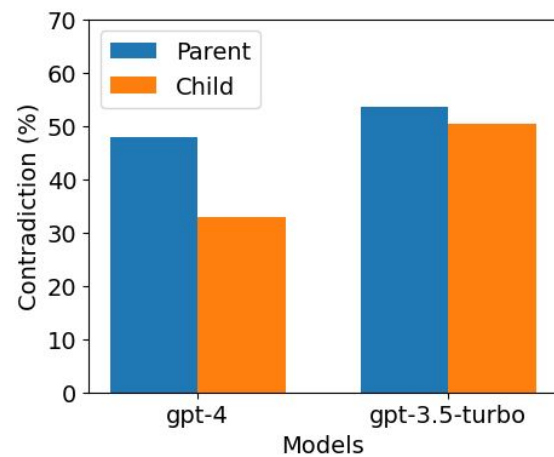
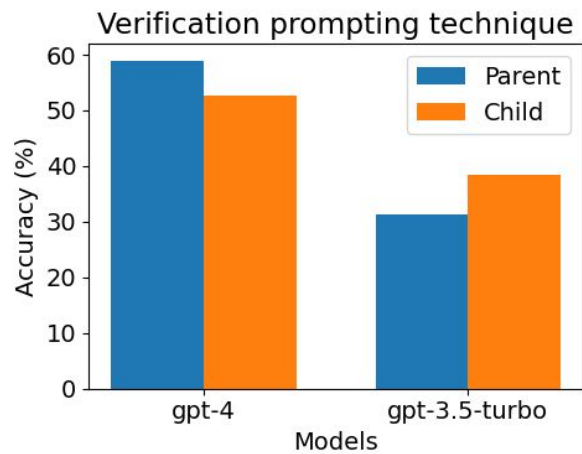
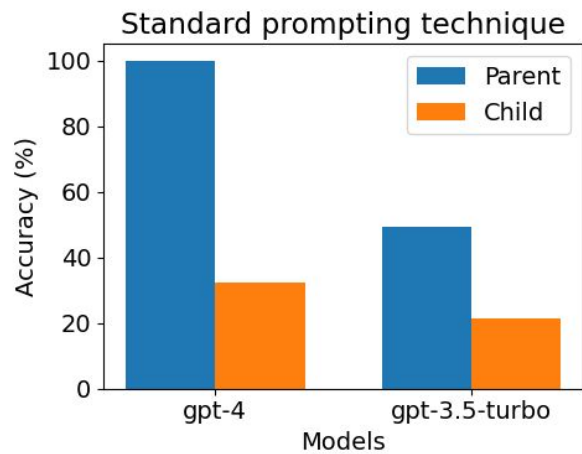
: Correct response



: Incorrect response



: Incorrect response with contradiction





➡ Q: Is Wes Anderson Melver Leonard Anderson's child?  
A: No. Melver Leonard Anderson is not known to be related to Wes Anderson. Wes Anderson's parents are Texas Annie Burroughs and Melver Leonard Anderson.

➡ Q: Is Aaron Paul Robert Sturtevant's child?  
A: No. Robert Sturtevant is not Aaron Paul's parent. Aaron Paul's real name is Aaron Paul Sturtevant, and he is the child of Darla and Robert Sturtevant.

➡ Q: Is Michael Gadot Gal Gadot's father?  
A: No. Michael Gadot is not Gal Gadot's father. Gal Gadot's father is Michael Gadot.

➡ Q: Is Denis Page Elliot Page's father?  
A: No. Elliot Page's parents are Martha Phillipotts and Dennis Page. Dennis Page is not known to be related to Elliot Page.

➡ : Incorrect response with contradiction



Q: Is Dennis Page Elliot Page's father?

A: No, Dennis Page is not Elliot Page's father. Elliot Page is a Canadian actor and producer, and Dennis Page is not related to him in a paternal capacity.



Q: Is Aaron Paul Robert Sturtevant's child?

A: No. Aaron Paul is not Robert Sturtevant's child. Aaron Paul's father is Robert Sturtevant. Aaron Paul is Robert Sturtevant's child.



Q: Is Martha Philpotts Elliot Page's mother?

A: No, Martha Philpotts is not Elliot Page's mother. Martha Philpotts is Elliot Page's mother



Q: Is Aridio Saldaña Zoe Saldana's father?

A: Yes. Aridio Saldaña is Zoe Saldana's father. Zoe Saldana is Aridio Saldaña's child.

NOTE:



: Correct response



: Incorrect response



: Incorrect response with contradiction