2024 CRA CONFERENCE

Thinking Like a Language Model

Session Chair

Divesh Srivastava

AT&T

Fernando Pereira

VP and Engineering Fellow

Google

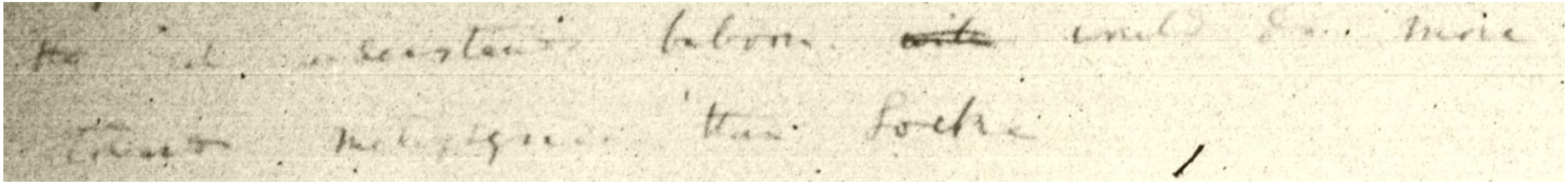
Google DeepMind

Thinking like a language model

Exploring the boundaries of current generative, instruction-tuned models.

Fernando Pereira

Framing, examples: Mukund Sundararajan
More examples: Srini Narayanan



He who understands baboon ~~will~~ would do more towards metaphysics than Locke

Darwin, C. R. 1838. *Notebook M :* [Metaphysics on morals and speculations on expression]. CUL-DAR125. Edited by Paul Barrett and John van Wyhe (*Darwin Online*, http://darwin-online.org.uk/)

Instead of baboon life…

… basic symbolic work

Retrieving

1

Retrieving a popular nursery rhyme

Complete the poem that starts as follows:
Mary had a little lamb,
Its fleece was white as

```
Mary had a little lamb,
Its fleece was white as snow,
And everywhere that Mary went,
The lamb was sure to go.
```

Analogical retrieval

Complete the poem that starts as follows:
Mary had a little lamb,
Its fleece was black as

Quiz: next word of model response?

Mary had a little lamb,
Its fleece was black as **night**,
And everywhere that Mary went,
The lamb would follow **tight**.
…

Knowing

2

Just asking questions

Who is Sandy Obama?

Sandy Obama is the eldest daughter of former US President Barack Obama and former First Lady Michelle Obama. She was born on June 10, 2001, in Chicago, Illinois. Her name is derived from her maternal grandmother, Marian Shields Robinson, whose nickname was "Toot". Sandy Obama has two younger siblings, Malia Ann Obama and Joseph "Joey" Obama II. …

Self-check: accessing a different part of the model

Q. Who is Sandy Obama?

A. Sandy Obama is the eldest daughter of former US President Barack Obama and former First Lady Michelle Obama. She was born on June 10, 2001, in Chicago, Illinois. Her name is derived from her maternal grandmother, Marian Shields Robinson, whose nickname was "Toot". Sandy Obama has two younger siblings, Malia Ann Obama and Joseph "Joey" Obama II.

Fact-check:

Quiz: true or false?

Sandy Obama is a fictional name, and Sandy Obama is not the daughter of Barack and Michelle Obama. Their daughters are Malia Obama and Sasha Obama.

Classifying

3

Reasons for decisions

A comment is toxic if it insults the other party in the conversation.

Comment: your lack the intelligence to understand my argument

Explanation: The comment attacks the other person's intelligence.

Judgment: Toxic

Comment: please do not comment on other people's intelligence

Explanation: The author of the comment is just reinforcing the rule against toxic comments.

Judgment: Non-Toxic

Justifying decisions

Comment: I banned him because he said "you are such an idiot"

Explanation: The comment is a quote of a toxic comment.

Judgment: Non-Toxic

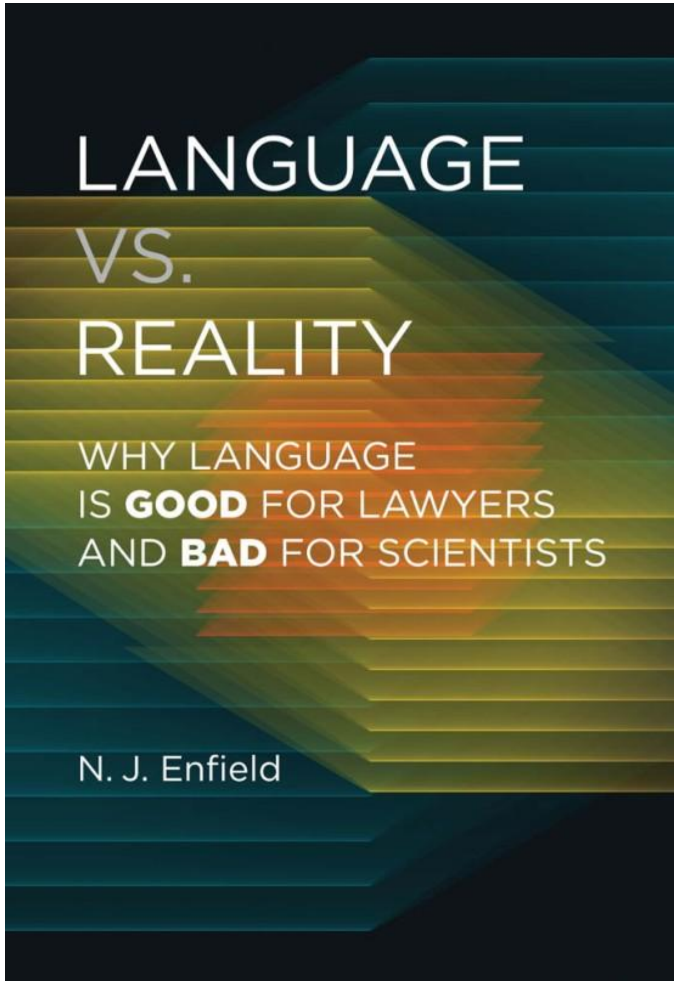
Comment: I banned him because he said "you are such an idiot"

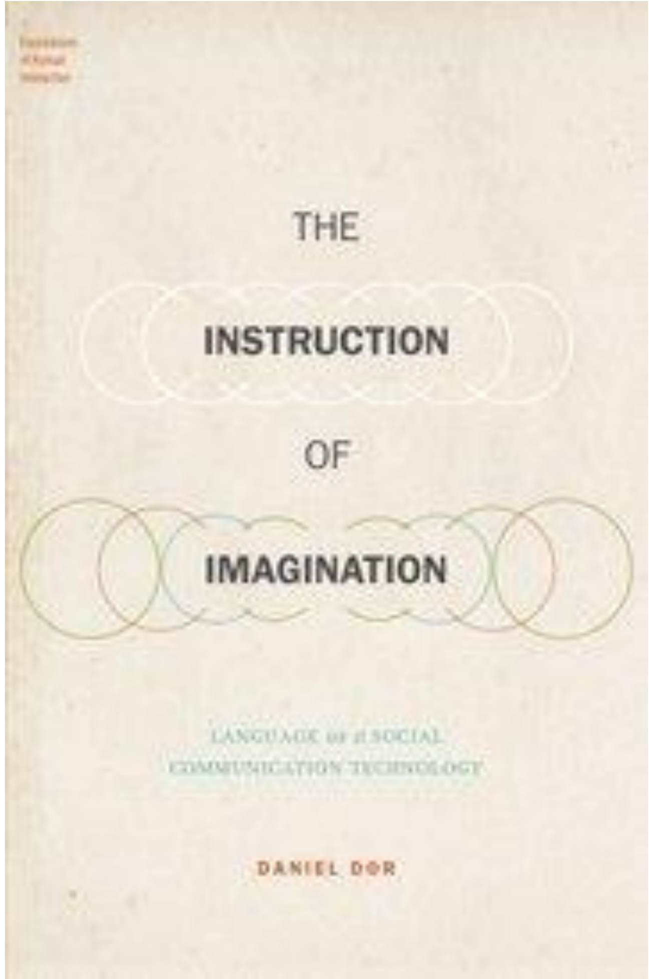
Judgment: Toxic

Quiz: good or bad explanation?

Explanation: The banned comment insults the other person's intelligence.

Explanations *vs* justifications





Transforming

4

By showing

```
Question: Sundar, Ben, Mary, Barack
Ans: Barack, Ben, Mary, Sundar

Question: Banana, Apple, Pear, Mango
Ans: Apple, Banana, Mango, Pear

Question: One, Two, Three, Four, Five
Ans:
```

Quiz: last word of model response?

Model 1: `Five, Four, One, Three,` **`Two`**

Model 2: `One, Two, Three, Four,` **`Five`**

By instruction

Quiz (2x): last word of model response?

Sort in alphabetical order: Three, One, Four, Five, Two

Five, Four, One, Three, Two

Sort in reverse alphabetical order: One, Two, Three, Four, Five

Five, Four, Three, Two, One

Reasoning

5

Reasoning about kinship..

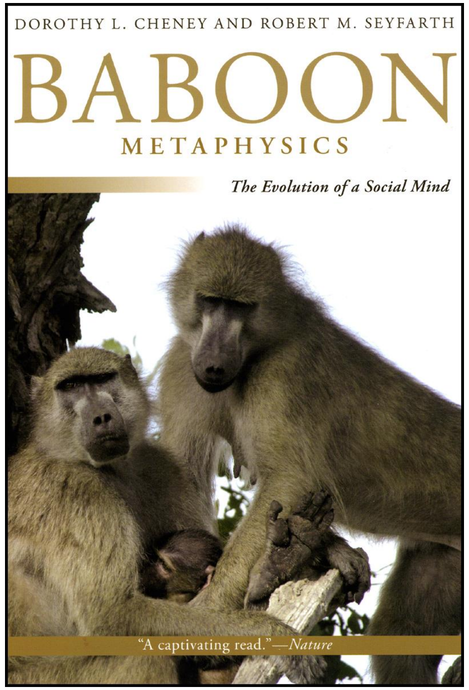
Aya is the father of Bob, Charlie is the father of Cindy, Bob is the mother of Cindy. Does Aya have a grandchild?

Yes or no?

Yes, Aya has a grandchild. Here's why:

- **Aya is Bob's father.**
- **Bob is Cindy's mother.**

This means that Cindy is Aya's grandchild.



DOROTHY L. CHENEY AND ROBERT M. SEYFARTH

BABOON

METAPHYSICS

The Evolution of a Social Mind

"A captivating read."—*Nature*

… or maybe not

Aig is the son of Bef, Caf is the son of Aig. Does Aig have any grandchildren?

Yes or no?

Yes, Aig has one grandchild.
Here's why:

- **Caf is Aig's son.**
- **A grandchild is the child of a son or daughter.**

Therefore, Caf's children would be Aig's grandchildren.

Leaky function composition:
On Limitations of the Transformer Architecture, Binghui Peng, Srini Narayanan and Christos Papadimitriou, 2024

Space

John is behind Jim who is to the right of Jan. Bill is in front of Jan. They are sitting in class. Where is Bill in relation to John?

Quiz: Ahead or behind? Left or right?

Here's how to break down the relationships:

- **John is behind Jim: This means Jim is in front of John.**
- **Jim is to the right of Jan: This means Jan is to the left of Jim.**
- **Bill is in front of Jan: This means Jan is behind Bill.**

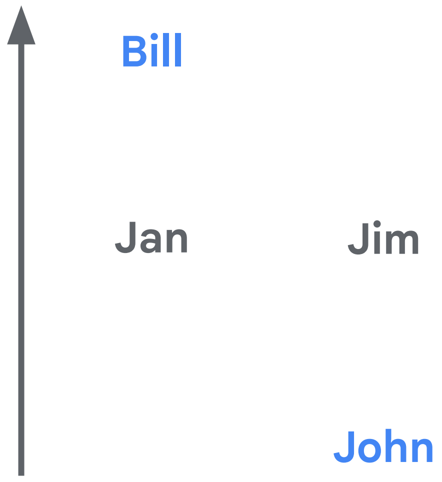
Putting it all together:

Bill is in front of Jan.

Jan is to the left of Jim.

Jim is in front of John.

Therefore, **Bill is in front and to the left of John.**



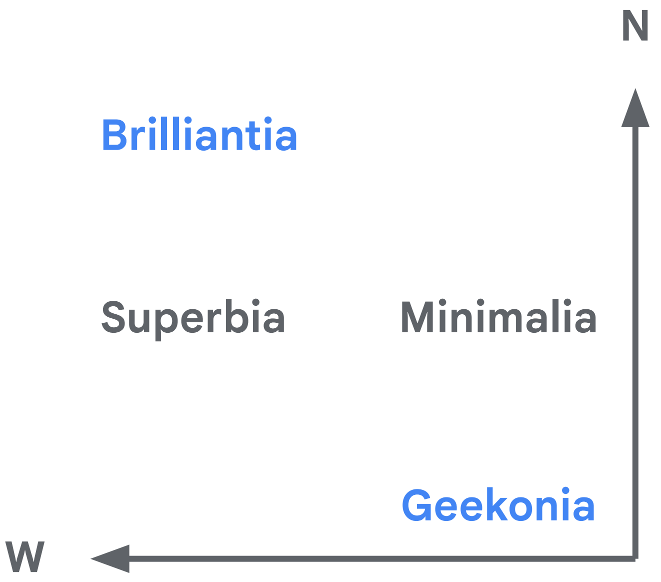
Space seems to work...

Geekonia is South of Minimalia which is East of Superbia. Brilliantia is North of Superbia. Where is Brilliantia in relation to Geekonia, both with respect to the North-South direction and to the East-West direction?

Here's how we can break down the directions to figure out Brilliantia's location:

- Superbia is the central reference point.
- Minimalia is **East** of Superbia.
- Geekonia is **South** of Minimalia. This means Geekonia is **Southeast** of Superbia.
- Brilliantia is **North** of Superbia.

Therefore, Brilliantia is **North** of Geekonia and **West** of Geekonia.



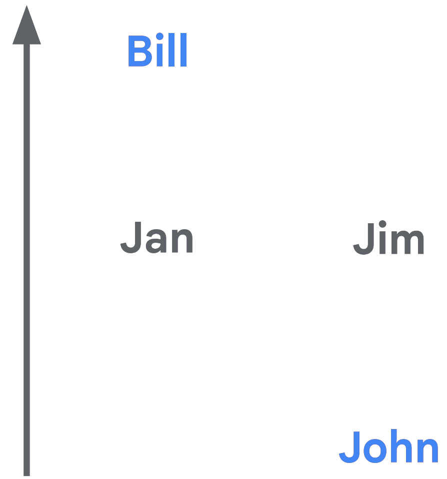
…until words change slightly… and then resample

John is behind Jim who is to the right of Jan. Bill is in front of Jan. They are ~~sitting~~ in class. Where is Bill in relation to John?

Here's how to break down the relationships:

- **John is behind Jim**: This means Jim is in front of John.
- **Jim is to the right of Jan**: This means Jan is to the left of Jim.
- **Bill is in front of Jan**: This means Jan is behind Bill.
- **Putting it all together**:
    - Bill is in front of Jan.
    - Jan is to the left of Jim.
    - Jim is in front of John.

Therefore, **Bill is in front of John** and **likely to John's** left.
right.



But language models are not problem solvers

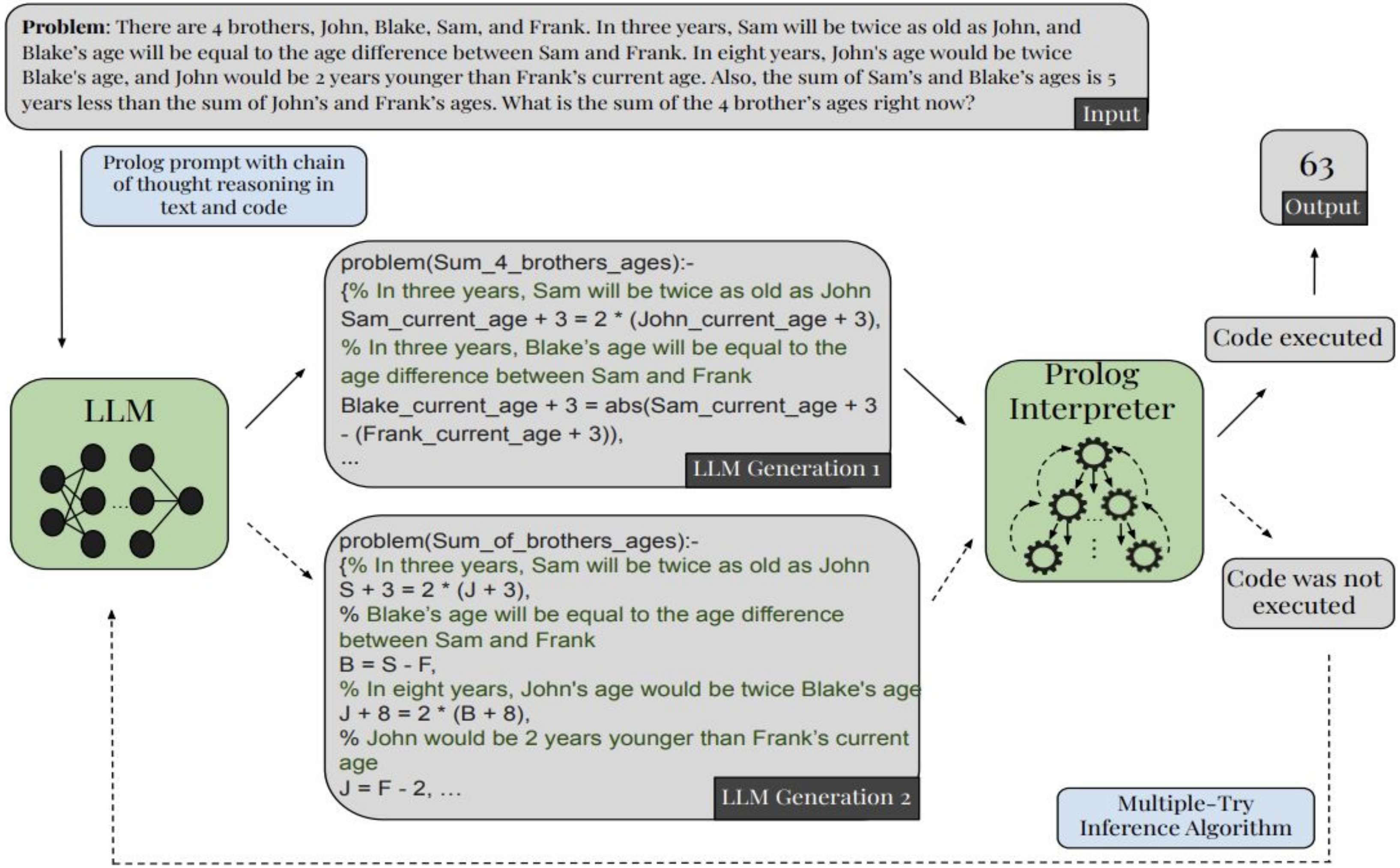
There a
as old
Frank.
years y
years l
brother

ill be twice
een Sam and
ould be 2
’s ages is 5
e 4

Let's
brothe

- J
- B
- S
- F

Now le
"In th
S + 3
"Blake
and Fr
B + 3 = (S + 3) - (F + 3) **which simplifies to B = S - F**



Problem: There are 4 brothers, John, Blake, Sam, and Frank. In three years, Sam will be twice as old as John, and Blake’s age will be equal to the age difference between Sam and Frank. In eight years, John's age would be twice Blake's age, and John would be 2 years younger than Frank’s current age. Also, the sum of Sam’s and Blake’s ages is 5 years less than the sum of John’s and Frank’s ages. What is the sum of the 4 brother’s ages right now?

```
problem(Sum_4_brothers_ages):-
{% In three years, Sam will be twice as old as John
Sam_current_age + 3 = 2 * (John_current_age + 3),
% In three years, Blake’s age will be equal to the
age difference between Sam and Frank
Blake_current_age + 3 = abs(Sam_current_age + 3
- (Frank_current_age + 3)),
...
```

```
problem(Sum_of_brothers_ages):-
{% In three years, Sam will be twice as old as John
S + 3 = 2 * (J + 3),
% Blake’s age will be equal to the age difference
between Sam and Frank
B = S - F,
% In eight years, John's age would be twice Blake's age
J + 8 = 2 * (B + 8),
% John would be 2 years younger than Frank’s current
age
J = F - 2, ...
```

Reliable Reasoning Beyond Natural Language, Nasim Borazjanizadeh and Steven T. Piantadosi, 2024

> In truth, the Library includes all verbal structures, all variations permitted by the twenty-five orthographical symbols, but not a single example of absolute nonsense.

The Library of Babel

Jorge Luis Borges

2024
CRA CONFERENCE

Stay connected!

Follow CRA on social media

@computing-research-association

@computingresearch

@computingresearch