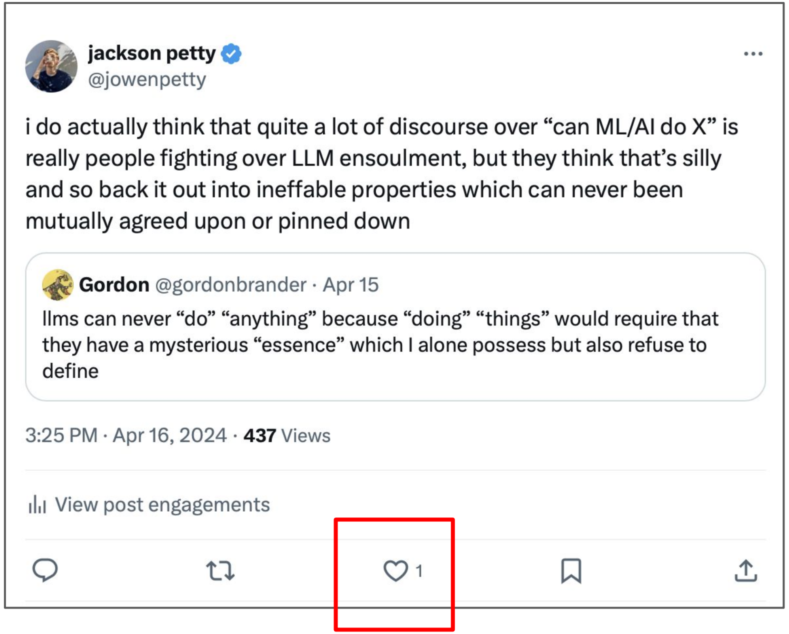
esoteric topics in language modeling

Can language models do _*anything*_?



jackson petty
@jowenpetty

i do actually think that quite a lot of discourse over “can ML/AI do X” is really people fighting over LLM ensoulment, but they think that’s silly and so back it out into ineffable properties which can never been mutually agreed upon or pinned down

Gordon @gordonbrander · Apr 15

llms can never “do” “anything” because “doing” “things” would require that they have a mysterious “essence” which I alone possess but also refuse to define

3:25 PM · Apr 16, 2024 · **437** Views

my hot take on this is that most people are actually arguing about something else entirely...

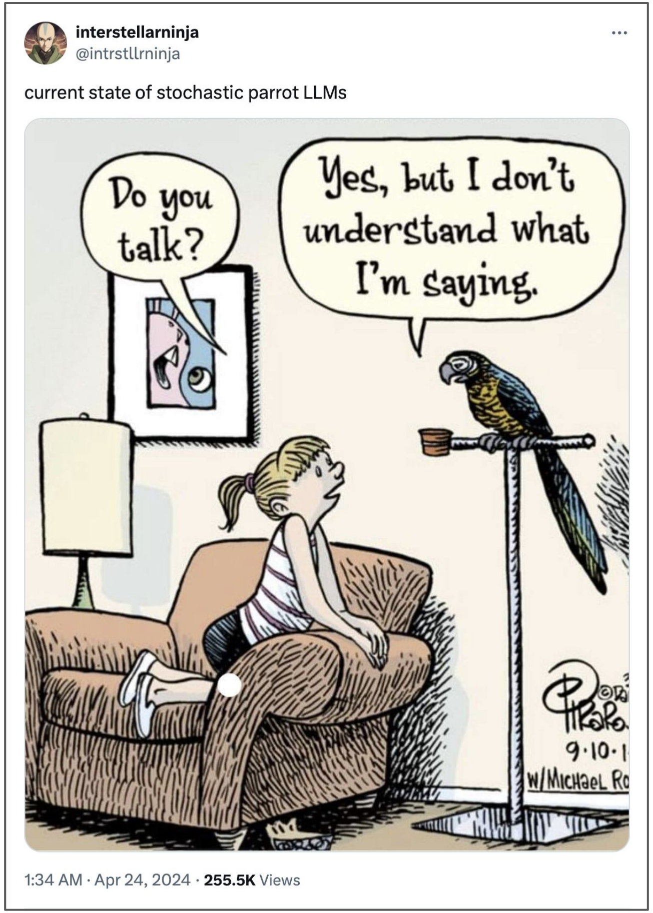
(as you can see, i am very popular on twitter)

Can language models *understand*?

are LLMs “just” “stochastic parrots”?

(term coined by Emily Bender et al.)

how can you tell? how could we tell if a parrot “understands” language? how can i tell if sophie understands language?



interstellarninja
@intrstllrninja

current state of stochastic parrot LLMs

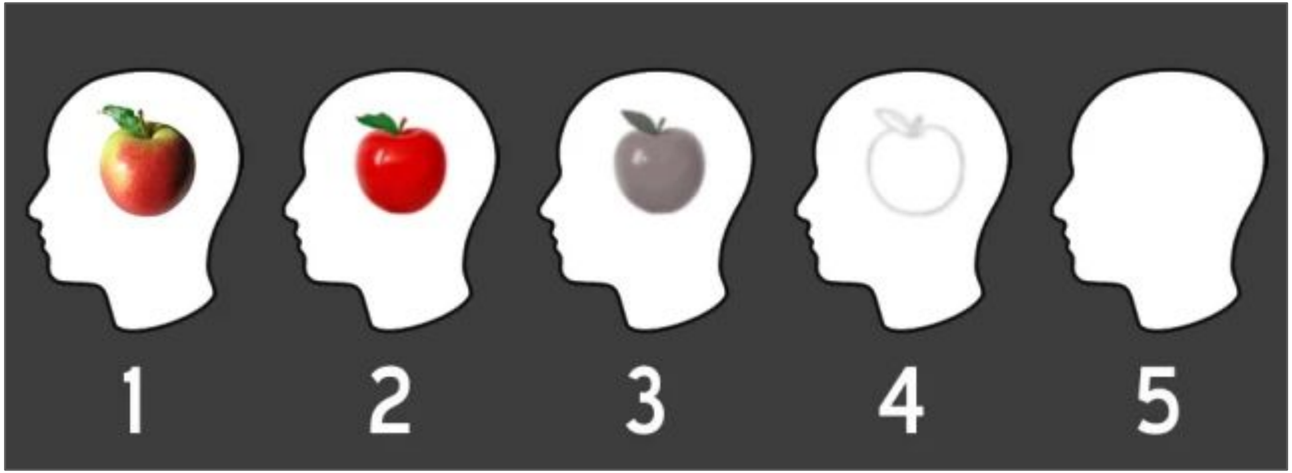
1:34 AM · Apr 24, 2024 · 255.5K Views

What does “understanding” mean?

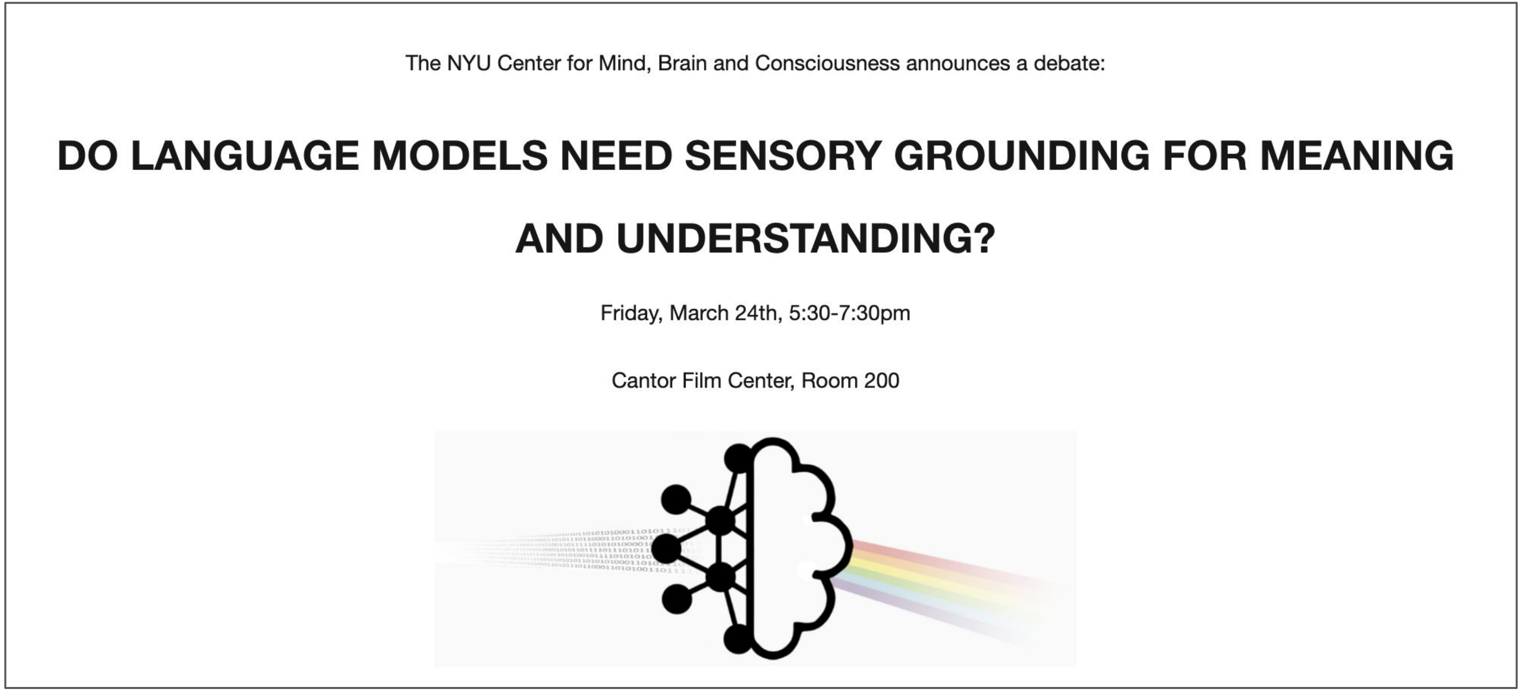
“Understanding” qua **behavior**: a model “understands” if it behaves as if it understands on some set of tasks

“Understanding” qua **experience**: a model “understands” if it experiences as if it understands

Sidequest(ion): *Berkeleyan idealism*



What is necessary to *understand*?



The NYU Center for Mind, Brain and Consciousness announces a debate:

DO LANGUAGE MODELS NEED SENSORY GROUNDING FOR MEANING AND UNDERSTANDING?

Friday, March 24th, 5:30-7:30pm

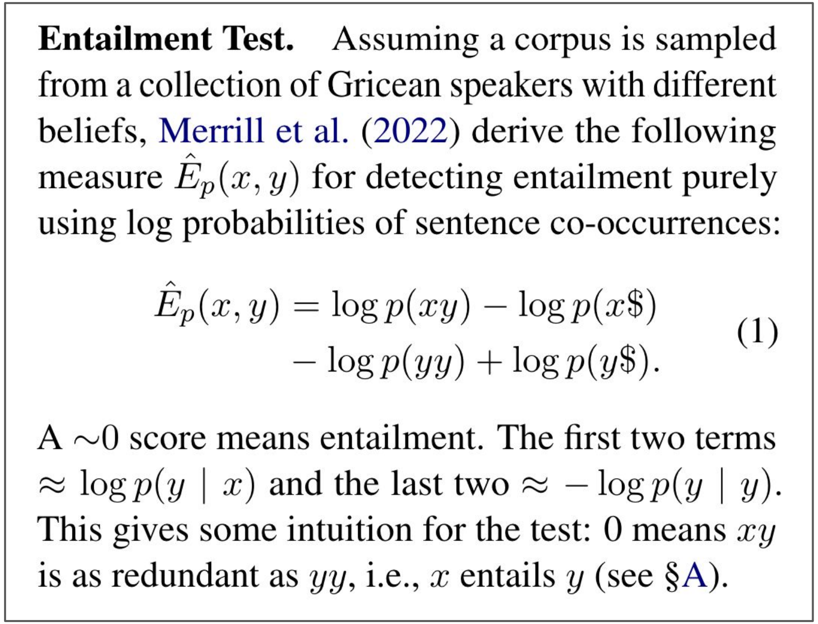
Cantor Film Center, Room 200

Ungrounded understanding

What does Mary know about the color "red"?

What happens when she goes outside?

Can LLMs learn semantics w/o grounding?



Entailment Test. Assuming a corpus is sampled from a collection of Gricean speakers with different beliefs, Merrill et al. (2022) derive the following measure $\hat{E}_p(x, y)$ for detecting entailment purely using log probabilities of sentence co-occurrences:

$$\begin{aligned} \hat{E}_p(x, y) = \log p(xy) - \log p(x\$) \\ - \log p(yy) + \log p(y\$). \end{aligned} \tag{1}$$

A ∼0 score means entailment. The first two terms $\approx \log p(y \mid x)$ and the last two $\approx - \log p(y \mid y)$. This gives some intuition for the test: 0 means xy is as redundant as yy, i.e., x entails y (see §A).

Pragmatics to the rescue!

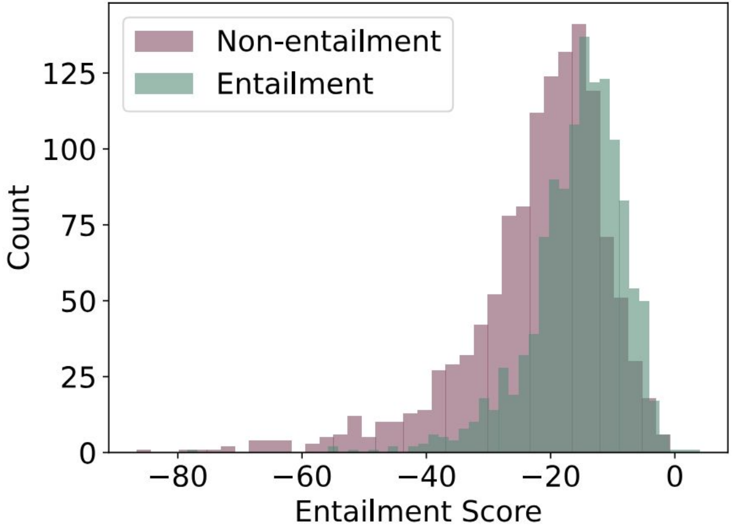


Figure 1: Entailment score $\hat{E}_p(x, y)$ distribution computed with Llama2-70b probabilities on RTE. **The score discriminates the two classes, though imperfectly.**

What does grounding give us, empirically?

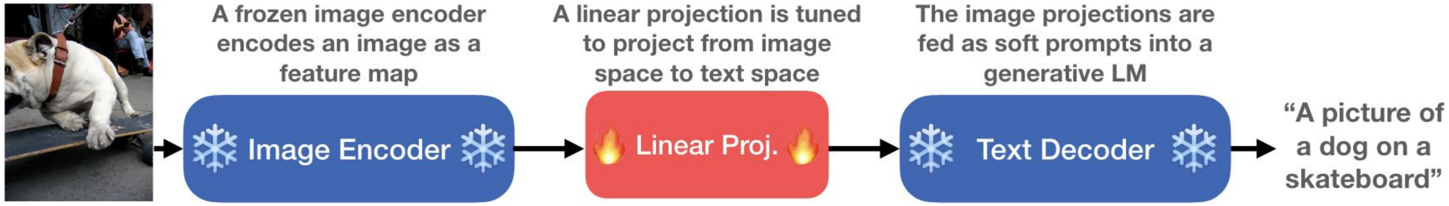
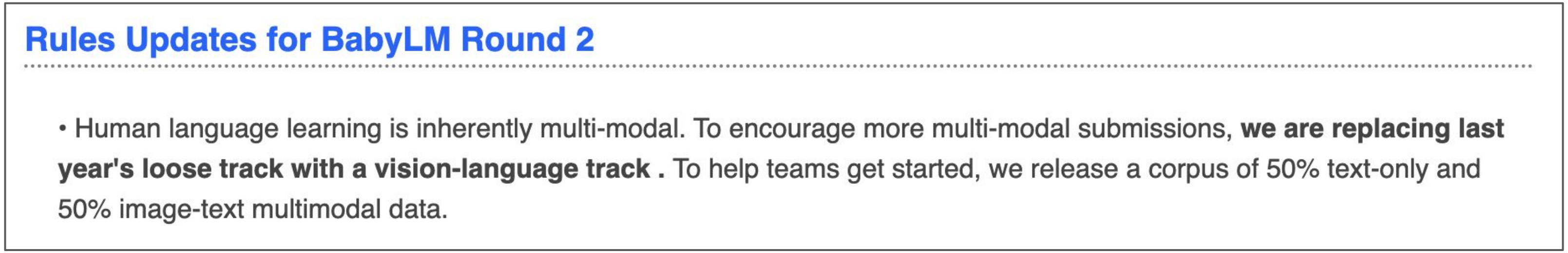

A frozen image encoder encodes an image as a feature map
A linear projection is tuned to project from image space to text space
The image projections are fed as soft prompts into a generative LM
Image Encoder
Linear Proj.
Text Decoder
"A picture of a dog on a skateboard"

Figure 1: We train linear projections from image representations into the input space of a language model to produce captions describing images. We find that LMs can describe the contents of most image representations, but performance varies based on the type of image encoder used.

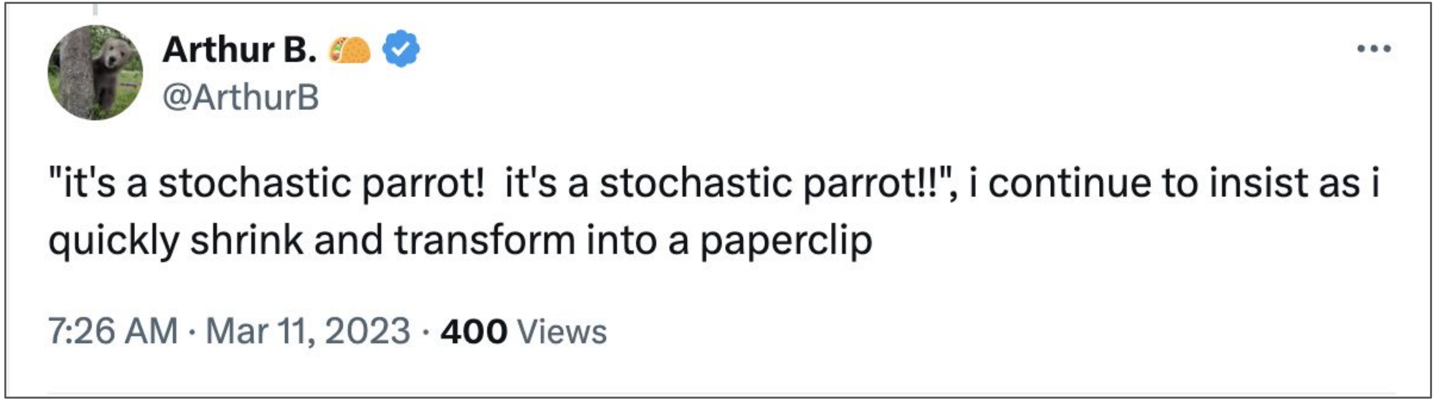
Merullo et al. (2023)



Rules Updates for BabyLM Round 2

- Human language learning is inherently multi-modal. To encourage more multi-modal submissions, **we are replacing last year's loose track with a vision-language track .** To help teams get started, we release a corpus of 50% text-only and 50% image-text multimodal data.

Does it *matter* if an LLM can ‘understand’?



Arthur B. 🌮 ✓
@ArthurB

"it's a stochastic parrot! it's a stochastic parrot!!", i continue to insist as i quickly shrink and transform into a paperclip

7:26 AM · Mar 11, 2023 · **400** Views