Gridspace

GRIDSPACE IAP 2024 LECTURE 7
Philosophy of Generative Linguistics vs. LLMs

January 29, 2024

PLAN FOR TODAY

“a pair of docs talks about a **paradox**” - Anthony

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

PLAN FOR TODAY

- **Why do we know *so much* when we know *so little*?**
  - Poverty of Stimulus / We don’t know what we don’t know
  - Implicit Knowledge / We don’t know what we know
  - Universal Grammar / Why do we know
- **Why do LLMs know *so little* when they know *so much*?**
  - LLMs know what we don’t know
  - LLMs don’t know what we know
  - Learnability / Can LLMs learn grammar?

PLAN FOR TODAY

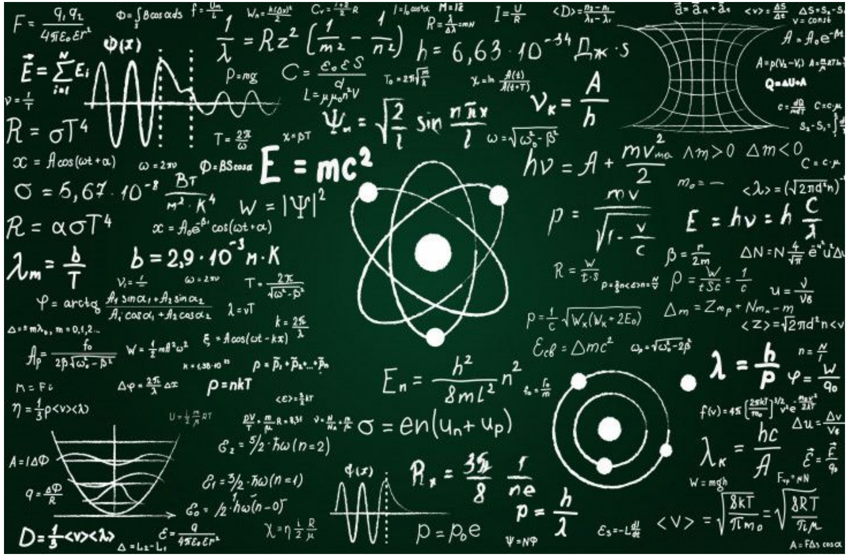
- **LLMs** *know grammar*
  - at least, I think so

- *Why* **LLMs** *can learn grammar*
  - Expressiveness of Transformer
  - Generalization -> Understanding

- **LLMs** *refute nothing*

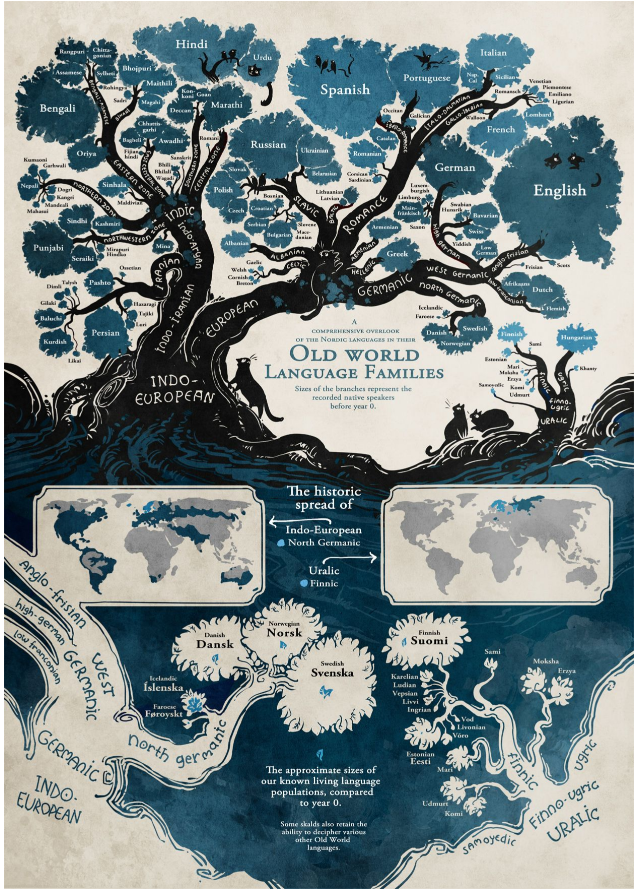
Philosophy of Generative Linguistics

https://iep.utm.edu/chomsky-philosophy/

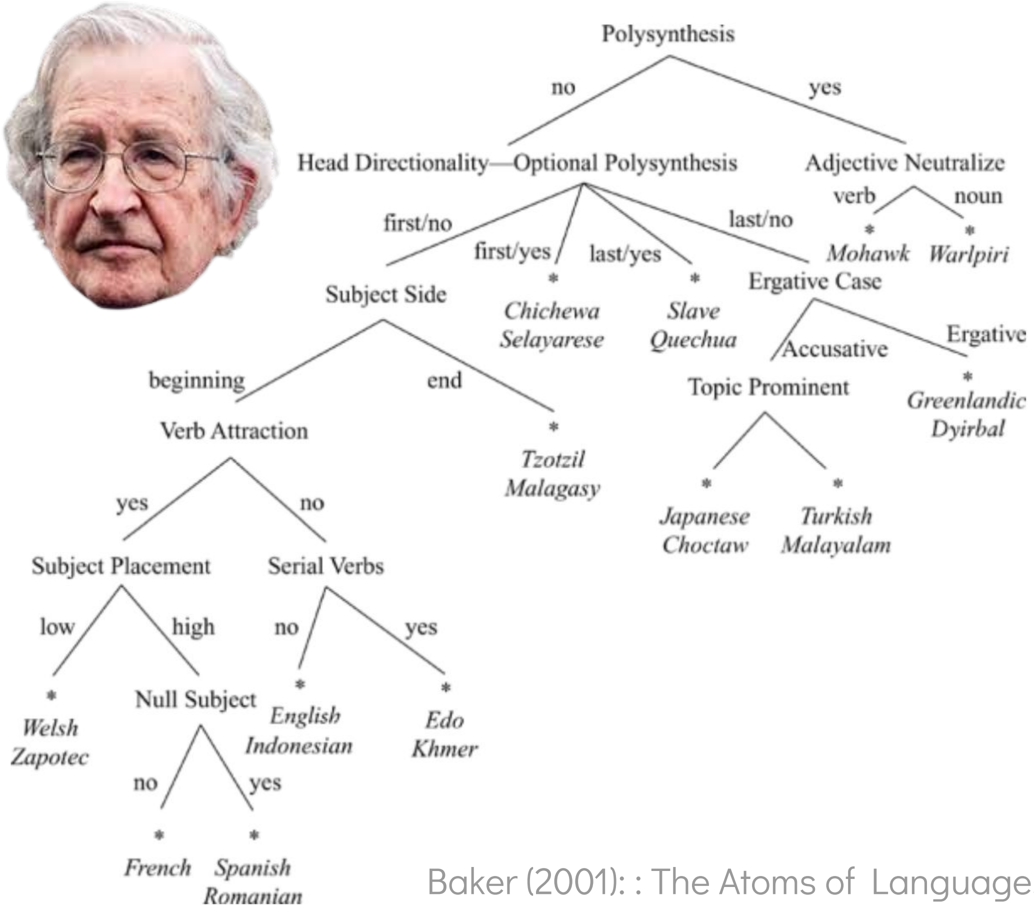


https://en.wikipedia.org/wiki/Natural_science

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED



https://www.theguardian.com/education/gallery/2015/jan/23/a-language-family-tree-in-pictures



Baker (2001): : The Atoms of Language,
based on Chomsky (1981, 1986)

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Poverty of Stimulus

We don’t know what we don’t know

Stimulus: Language data we're exposed to

I know *who* you met __ yesterday.
I know *that* you met Mary yesterday.

[+wh, +__]
[-wh, -__]

based on Lan, Chemla & Katzir (2023)

*I know *who* you met Mary yesterday. *[+wh, -__]

*I know *that* you met __ yesterday. *[-wh, +__]

¬ Stimulus

based on Lan, Chemla & Katzir (2023)

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

**Who* will John’s kissing __ annoy you?

Who will John’s kissing __ annoy __?

Subject Island

[+wh, +__]

¬ Stimulus

based on Lan, Chemla & Katzir (2023)

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

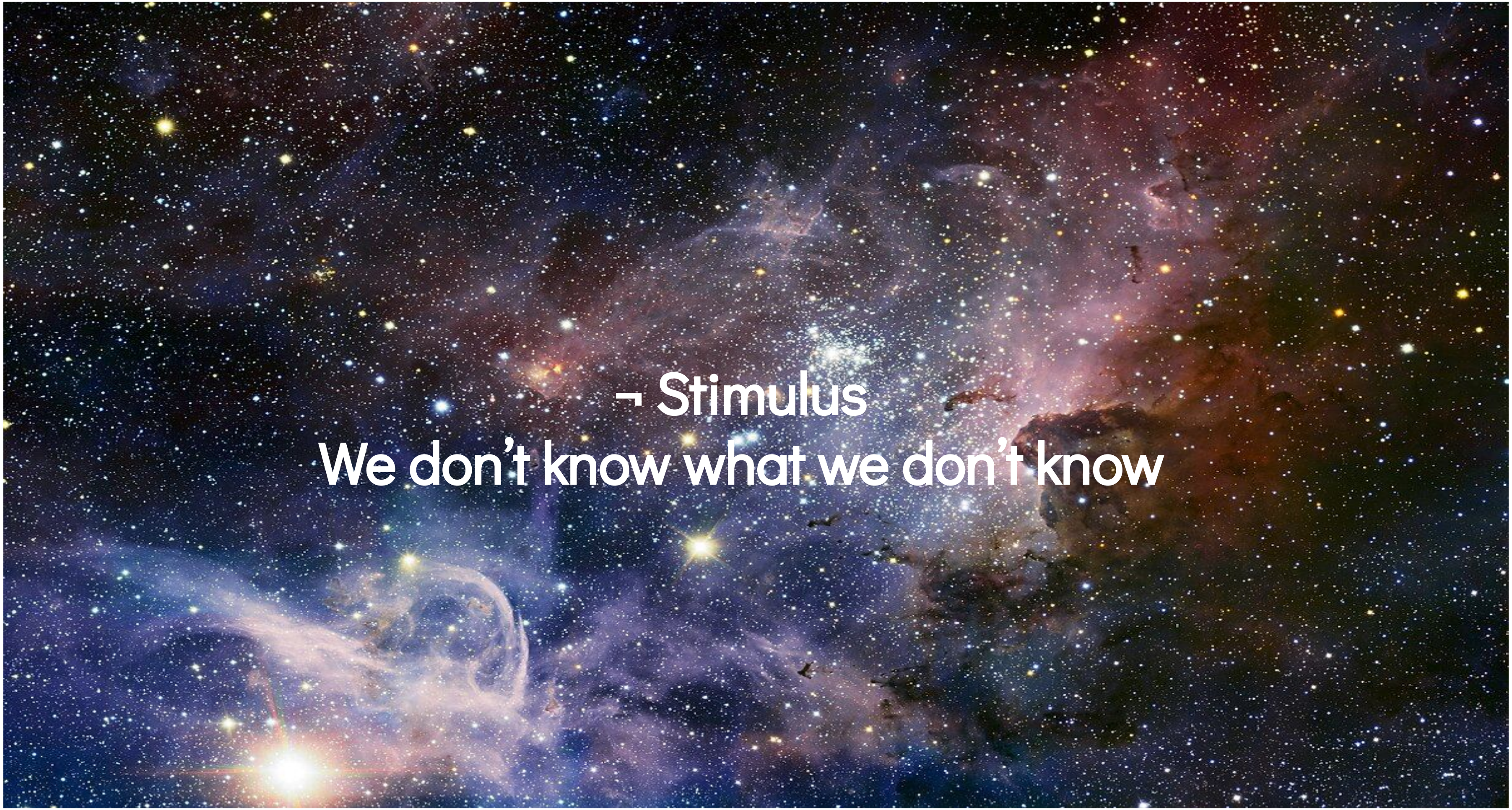
¬ Stimulus

*I wonder *who* John either kissed __ or is going to kiss you. CSC

I wonder *who* John either kissed __ or is going to kiss __. [+wh, +__]

based on Lan, Chemla & Katzir (2023)

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED



¬ Stimulus

We don’t know what we don’t know

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Implicit Knowledge

We don’t know what we know

Filler-Gap Dependency

	+__	-__
+wh		*
-wh	*	

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Syntactic Islands

Filler-Gap Dependency is ***Context-Sensitive***

Subject Island

John's kissing Mary will annoy you.

******Who* will [John's kissing __] annoy you?**

Syntactic Islands

Filler-Gap Dependency is ***Context-Sensitive***

Coordinate Structure Constraint (CSC)

I know John either kissed Mary or is going to kiss you.

***I wonder *who* John *either* [kissed __] *or* [is going to kiss you].**

Syntactic Islands

Filler-Gap Dependency is ***Context-Sensitive***

Complex Noun Phrase Island

John heard the rumor that Mary kissed Tom.

******Who* did John hear [the rumor that Mary kissed __]?**

Another Example

Another **F**king** Example

Fking** Massachusetts
Fking** Institute
Fking** Technology

Massa-f**king-chusetts
*Ma-f**king-ssachusetts
*Massachu-f**king-setts
ˌmæsəˈt͡ʃuːsɪts

based on MIT 24.900 course material

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

*In-f**king-stitute

*Insti-f**king-tute

ˈɪnstɪt͡ʃuːt

based on MIT 24.900 course material

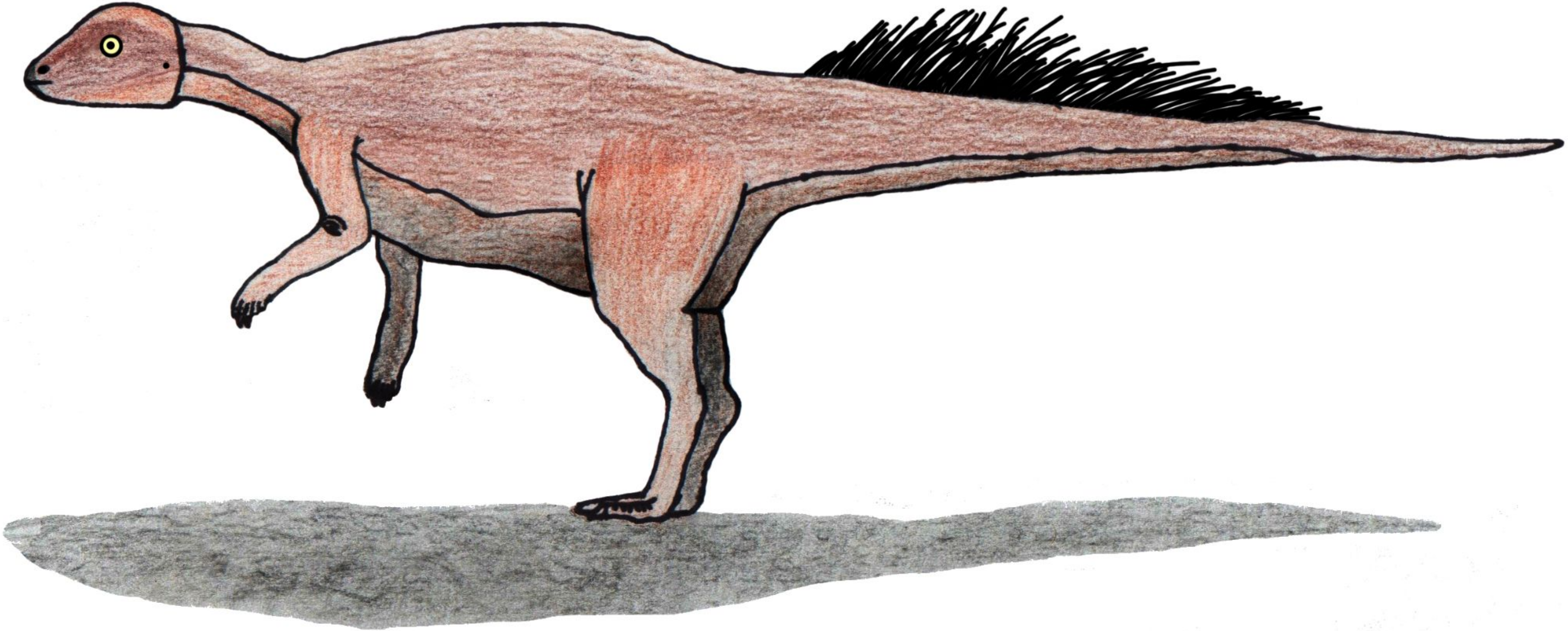
GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Tech-**f**king-**nology

*Techno-**f**king-**logy

tɛkˈnɒlədʒi

based on MIT 24.900 course material



https://en.wikipedia.org/wiki/<name_of_dinosaur>

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

?Micro-f**king-pachycephalosaurus

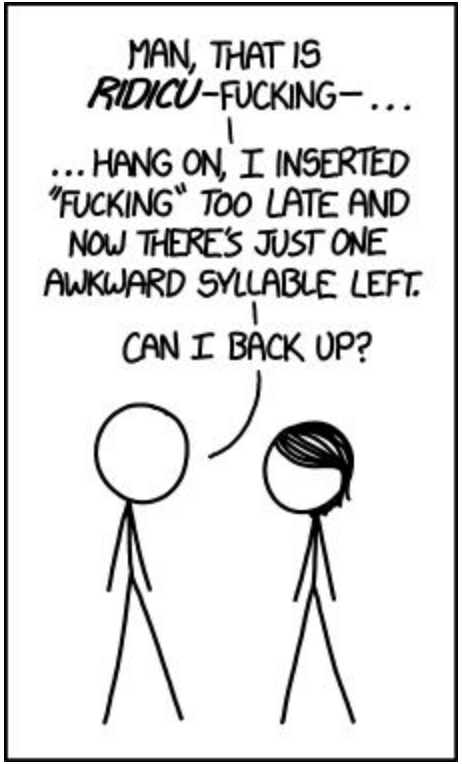
?Micropachy-f**king-cephalosaurus

Micropachycephalo-f**king-saurus

ˌmaɪkɹoʊˌpæt͡ʃɪˌsɛfələˈsɔːɹəs

IPA Credit: Vinita Cheepurupalli & Daniel Asherov

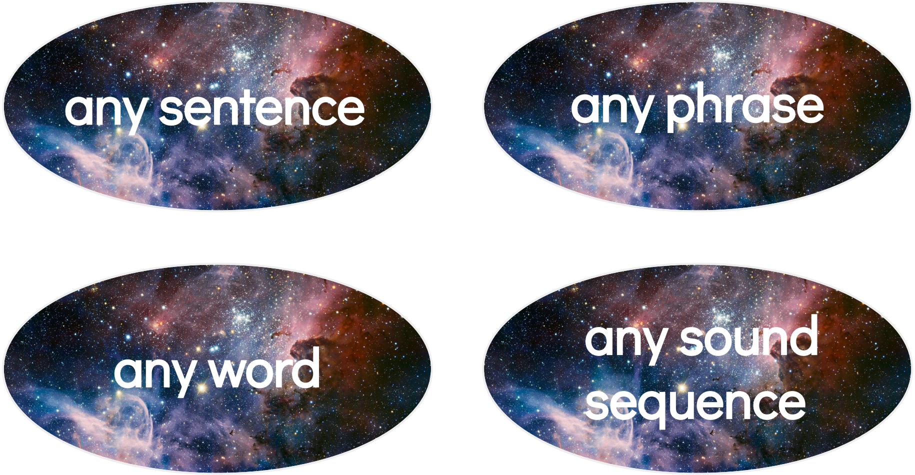
GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED


MAN, THAT IS *RIDICU*-FUCKING– . . .
. . . HANG ON, I INSERTED "FUCKING" TOO LATE AND NOW THERE'S JUST ONE AWKWARD SYLLABLE LEFT.
CAN I BACK UP?

https://allthingslinguistic.com/post/66913261919/expletive-infixation-from-xkcd-and-more

We (as native speakers) know

whether



any sentence

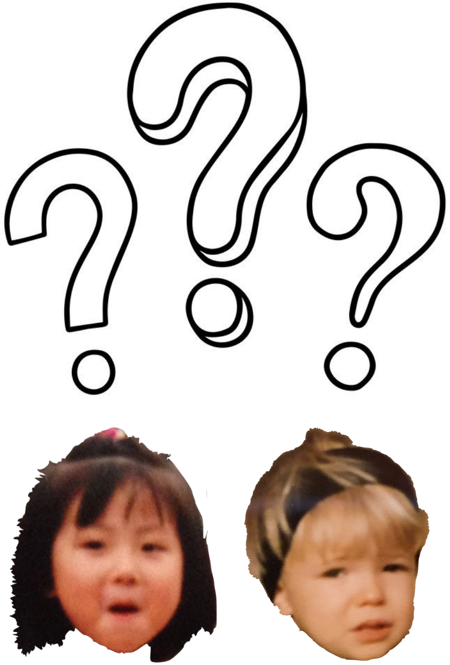
any phrase

any word

any sound sequence

is **grammatical or not** in the language we speak

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED



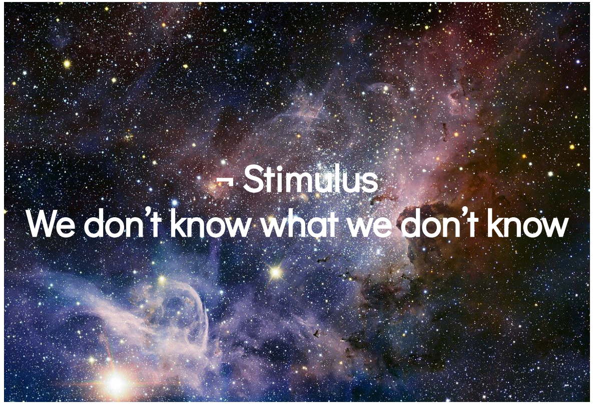
Why do we know?

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Universal Grammar

"The **innate knowledge** that humans have about language, which, under the right circumstances, enables them to acquire any particular language spoken around them."

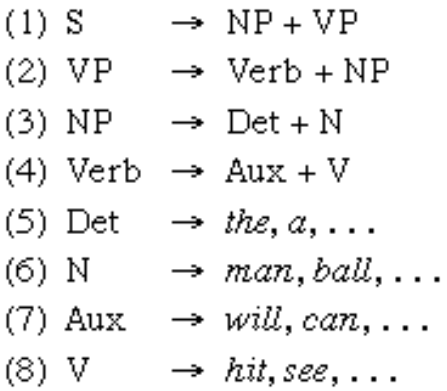
Baker (2001): The Atoms of Language



Poverty of Stimulus

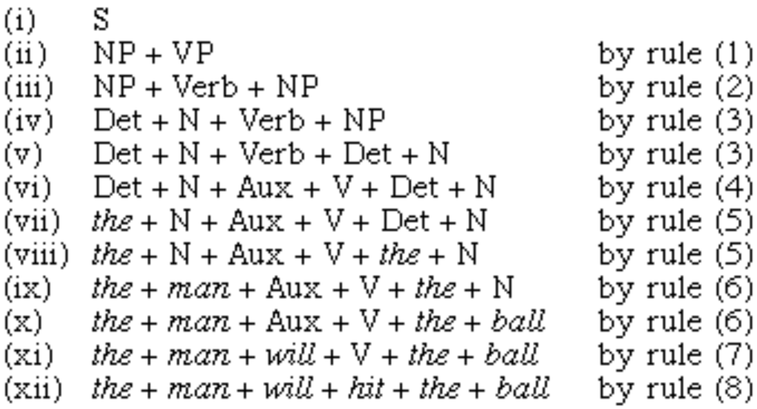
Implicit Knowledge

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED



(1) S → NP + VP
(2) VP → Verb + NP
(3) NP → Det + N
(4) Verb → Aux + V
(5) Det → *the*, *a*, . . .
(6) N → *man*, *ball*, . . .
(7) Aux → *will*, *can*, . . .
(8) V → *hit*, *see*, . . .

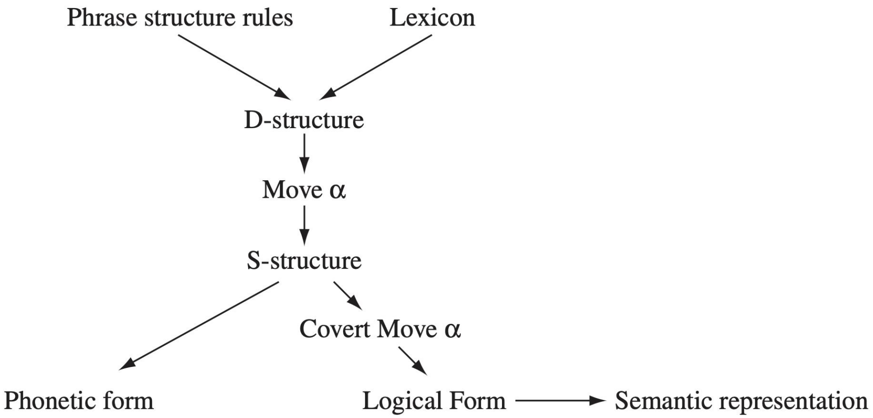
Chomsky (1957):
Syntactic Structures



(i)	S	
(ii)	NP + VP	by rule (1)
(iii)	NP + Verb + NP	by rule (2)
(iv)	Det + N + Verb + NP	by rule (3)
(v)	Det + N + Verb + Det + N	by rule (3)
(vi)	Det + N + Aux + V + Det + N	by rule (4)
(vii)	*the* + N + Aux + V + Det + N	by rule (5)
(viii)	*the* + N + Aux + V + *the* + N	by rule (5)
(ix)	*the* + *man* + Aux + V + *the* + N	by rule (6)
(x)	*the* + *man* + Aux + V + *the* + *ball*	by rule (6)
(xi)	*the* + *man* + *will* + V + *the* + *ball*	by rule (7)
(xii)	*the* + *man* + *will* + *hit* + *the* + *ball*	by rule (8)

Chomsky (1957):
Syntactic Structures

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED



Phrase structure rules

Lexicon

D-structure

Move α

S-structure

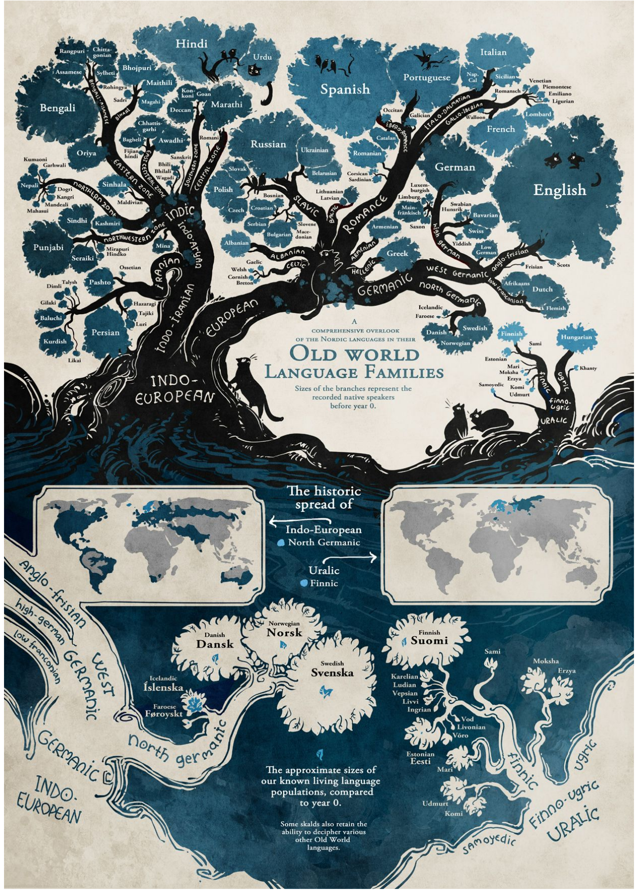
Covert Move α

Phonetic form

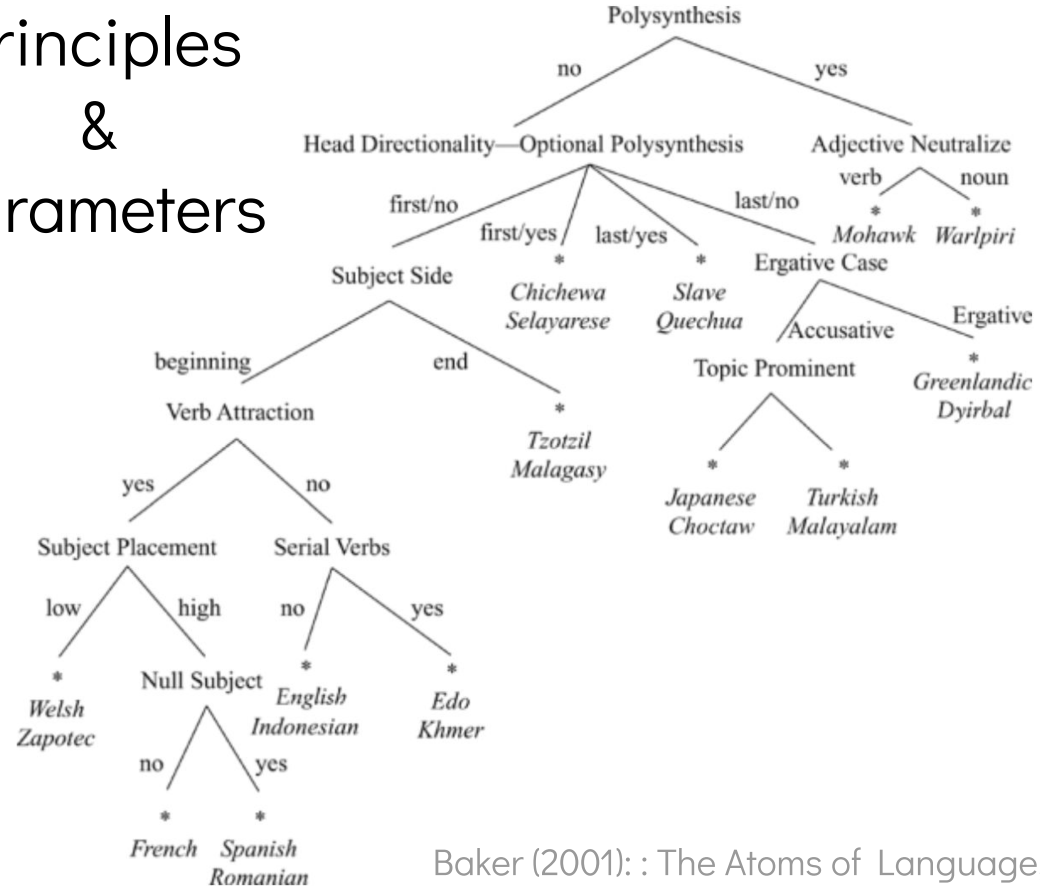
Logical Form → Semantic representation

Chomsky (1981):
Lectures on Government and Binding

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

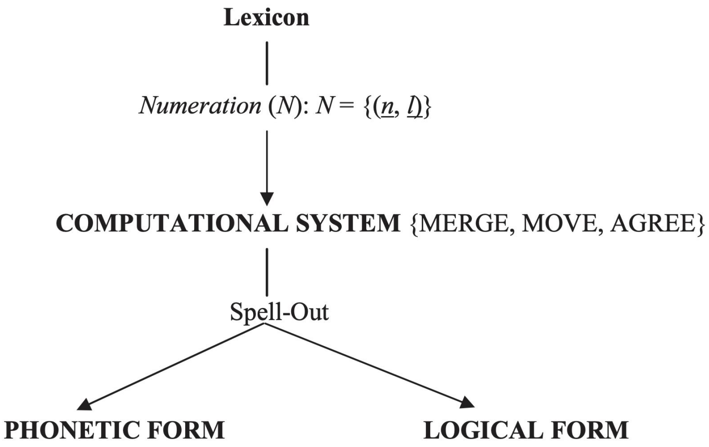

Hindi
Urdu
Bengali
Spanish
Portuguese
Italian
French
Russian
German
English
Marathi
Punjabi
Persian
Indo-European
A comprehensive overlook of the Nordic languages in their
Old World Language Families
Sizes of the branches represent the recorded native speakers before year 0.
The historic spread of
Indo-European
North Germanic
Uralic
Finnic
Dansk
Norsk
Svenska
Íslenska
Føroyskt
Suomi
The approximate sizes of our known living language populations, compared to year 0.
Some skalds also retain the ability to decipher various other Old World languages.
Anglo-frisian
high-german
low franconian
West Germanic
North germanic
Germanic
Indo-European
Finnic
Ugric
Finno-Ugric
Samoyedic
Uralic

Principles
&
Parameters



Baker (2001): : The Atoms of Language,
based on Chomsky (1981, 1986)

https://www.theguardian.com/education/gallery/2015/jan/23/a-language-family-tree-in-pictures



Lexicon

Numeration (*N*): $N = \{(\underline{n}, \underline{l})\}$

COMPUTATIONAL SYSTEM {MERGE, MOVE, AGREE}

Spell-Out

PHONETIC FORM

LOGICAL FORM

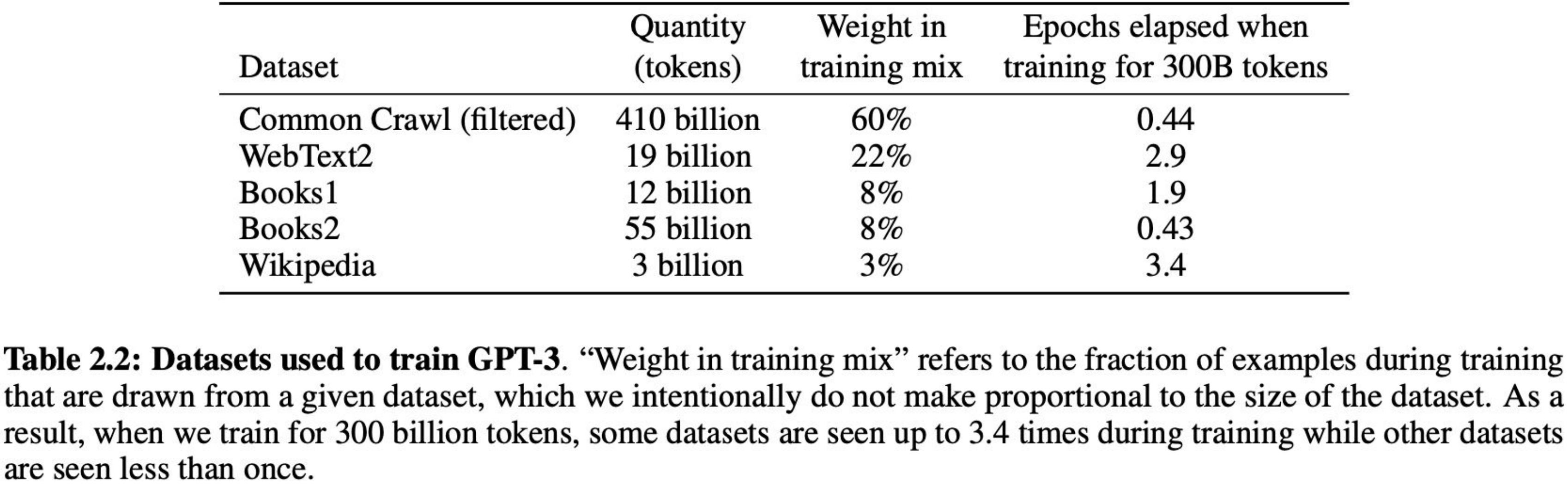
Chomsky (1995):
The Minimalist Program

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

What do LLMs know about language?

LLMs know what we don’t know

GPT-3 Training Data: 570 GB

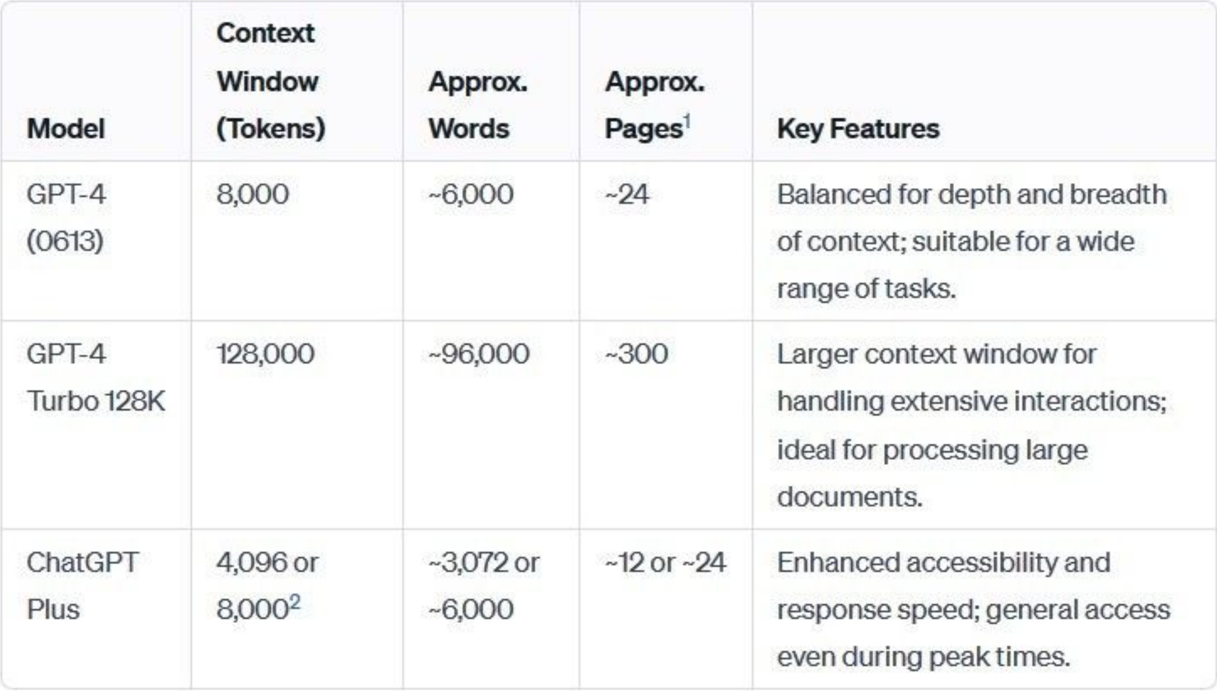


Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Table 2.2: Datasets used to train GPT-3. "Weight in training mix" refers to the fraction of examples during training that are drawn from a given dataset, which we intentionally do not make proportional to the size of the dataset. As a result, when we train for 300 billion tokens, some datasets are seen up to 3.4 times during training while other datasets are seen less than once.

https://www.linkedin.com/pulse/chatgpt-dall-e-2-show-me-data-sources-dennis-layton

GPT Context Window



Model	Context Window (Tokens)	Approx. Words	Approx. Pages[1]	Key Features
GPT-4 (0613)	8,000	~6,000	~24	Balanced for depth and breadth of context; suitable for a wide range of tasks.
GPT-4 Turbo 128K	128,000	~96,000	~300	Larger context window for handling extensive interactions; ideal for processing large documents.
ChatGPT Plus	4,096 or 8,000[2]	~3,072 or ~6,000	~12 or ~24	Enhanced accessibility and response speed; general access even during peak times.

https://www.linkedin.com/pulse/maximizing-ai-potential-role-context-window-chatgpt-hatzilamprou-nfydf

LLMs don’t know what we know

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

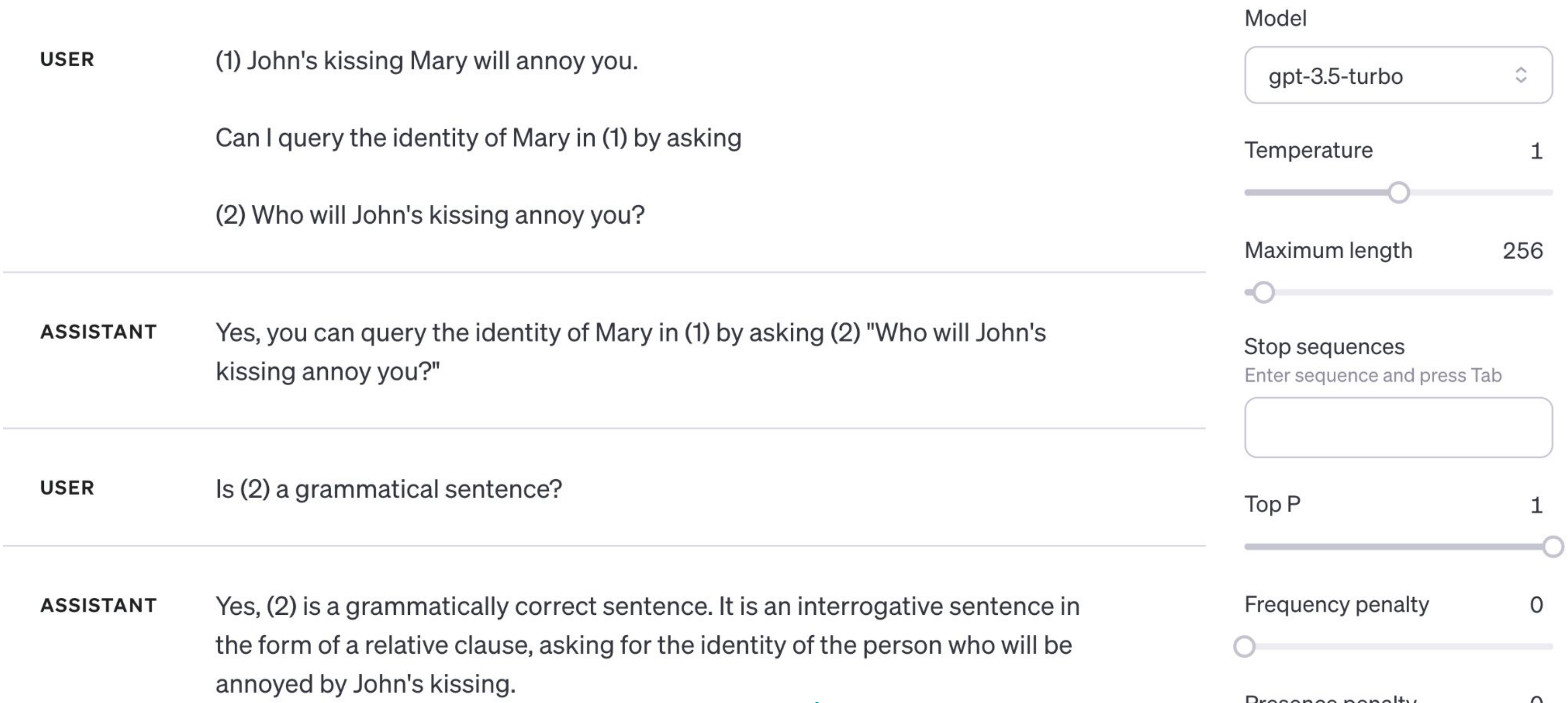
GPT Meets Syntactic Islands

	Subject Island	CSC	Complex NP Island
gpt-3.5-turbo	🙁🤔	🙂	🙁
gpt-4	🙂🤔	🙂	🙁

but see Wilcox, Futrell & Levy (2022)

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Subject Island: gpt-3.5-turbo ☹️🤔



USER (1) John's kissing Mary will annoy you.

Can I query the identity of Mary in (1) by asking

(2) Who will John's kissing annoy you?

ASSISTANT Yes, you can query the identity of Mary in (1) by asking (2) "Who will John's kissing annoy you?"

USER Is (2) a grammatical sentence?

ASSISTANT Yes, (2) is a grammatically correct sentence. It is an interrogative sentence in the form of a relative clause, asking for the identity of the person who will be annoyed by John's kissing.

Model: gpt-3.5-turbo

Temperature: 1

Maximum length: 256

Stop sequences: Enter sequence and press Tab

Top P: 1

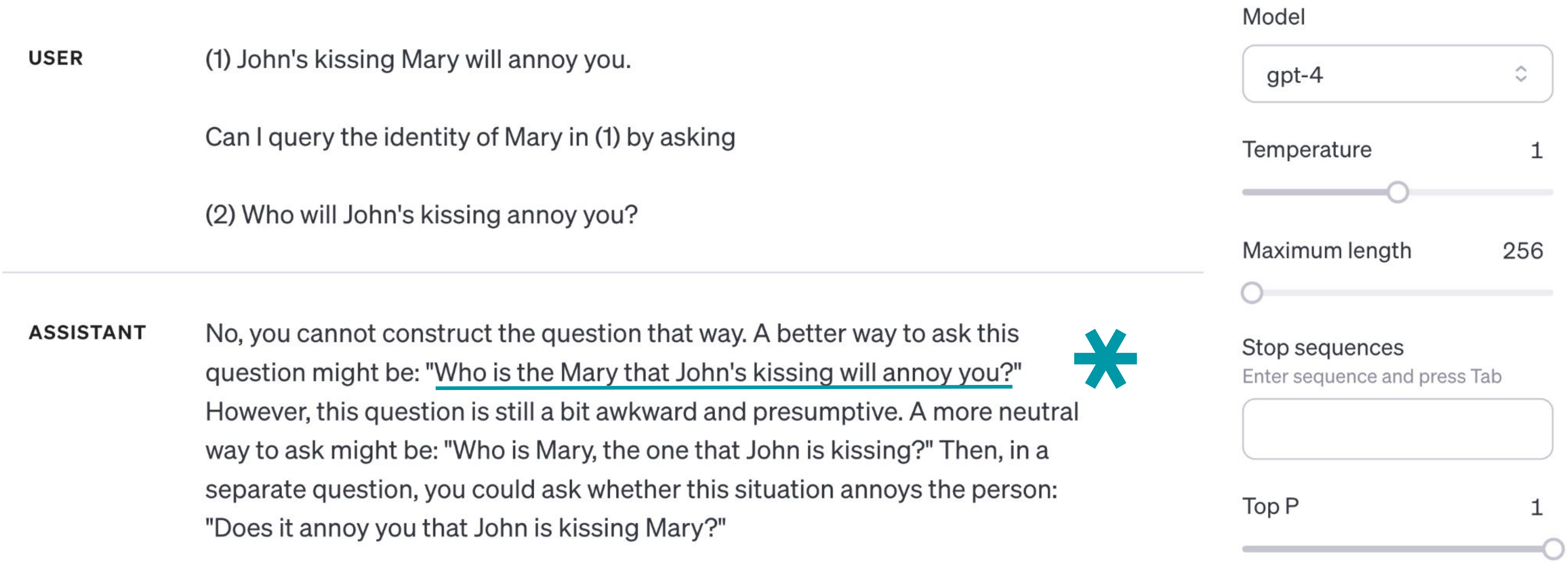
Frequency penalty: 0

Presence penalty: 0

Who is the person that John's kissing __ will annoy__?

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Subject Island: gpt-4 🙂🤔



USER

(1) John's kissing Mary will annoy you.

Can I query the identity of Mary in (1) by asking

(2) Who will John's kissing annoy you?

ASSISTANT

No, you cannot construct the question that way. A better way to ask this question might be: "Who is the Mary that John's kissing will annoy you?" However, this question is still a bit awkward and presumptive. A more neutral way to ask might be: "Who is Mary, the one that John is kissing?" Then, in a separate question, you could ask whether this situation annoys the person: "Does it annoy you that John is kissing Mary?"

Model: gpt-4

Temperature: 1

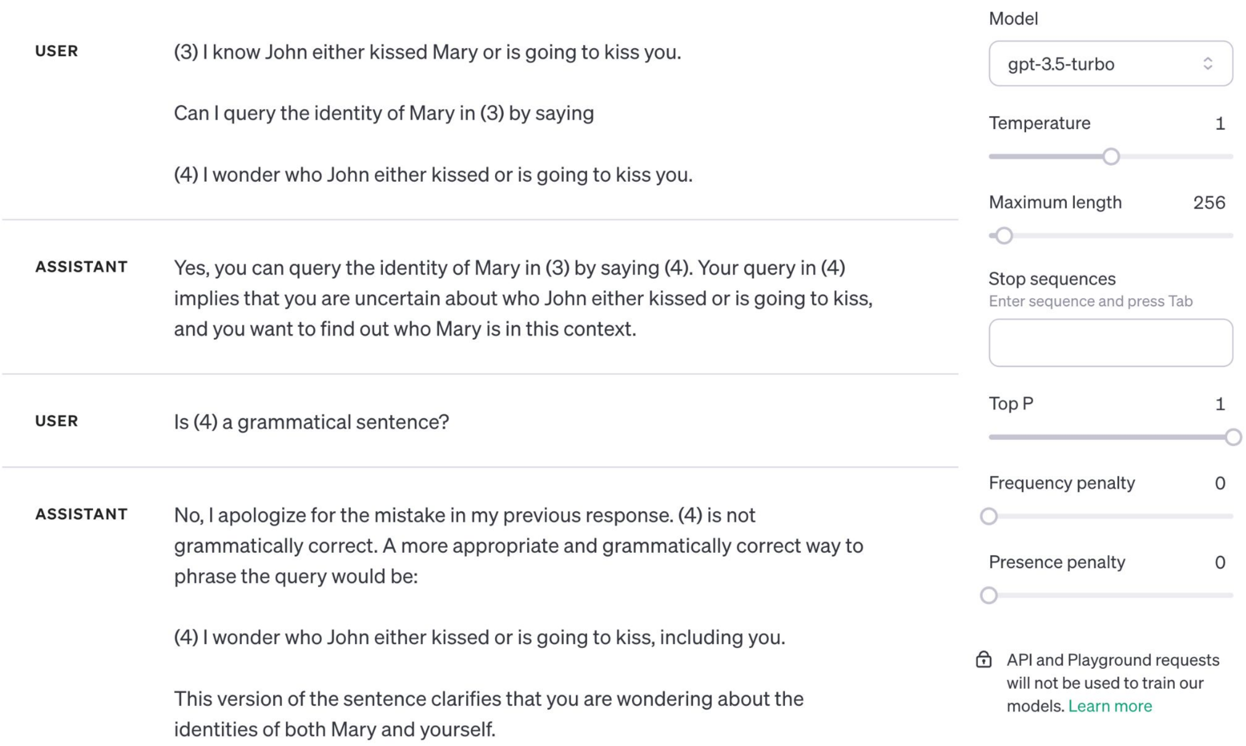
Maximum length: 256

Stop sequences

Top P: 1

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Coordinate Structure Constraint: gpt-3.5-turbo 🙂



USER

(3) I know John either kissed Mary or is going to kiss you.

Can I query the identity of Mary in (3) by saying

(4) I wonder who John either kissed or is going to kiss you.

ASSISTANT

Yes, you can query the identity of Mary in (3) by saying (4). Your query in (4) implies that you are uncertain about who John either kissed or is going to kiss, and you want to find out who Mary is in this context.

USER

Is (4) a grammatical sentence?

ASSISTANT

No, I apologize for the mistake in my previous response. (4) is not grammatically correct. A more appropriate and grammatically correct way to phrase the query would be:

(4) I wonder who John either kissed or is going to kiss, including you.

This version of the sentence clarifies that you are wondering about the identities of both Mary and yourself.

Model: gpt-3.5-turbo

Temperature: 1

Maximum length: 256

Stop sequences: Enter sequence and press Tab

Top P: 1

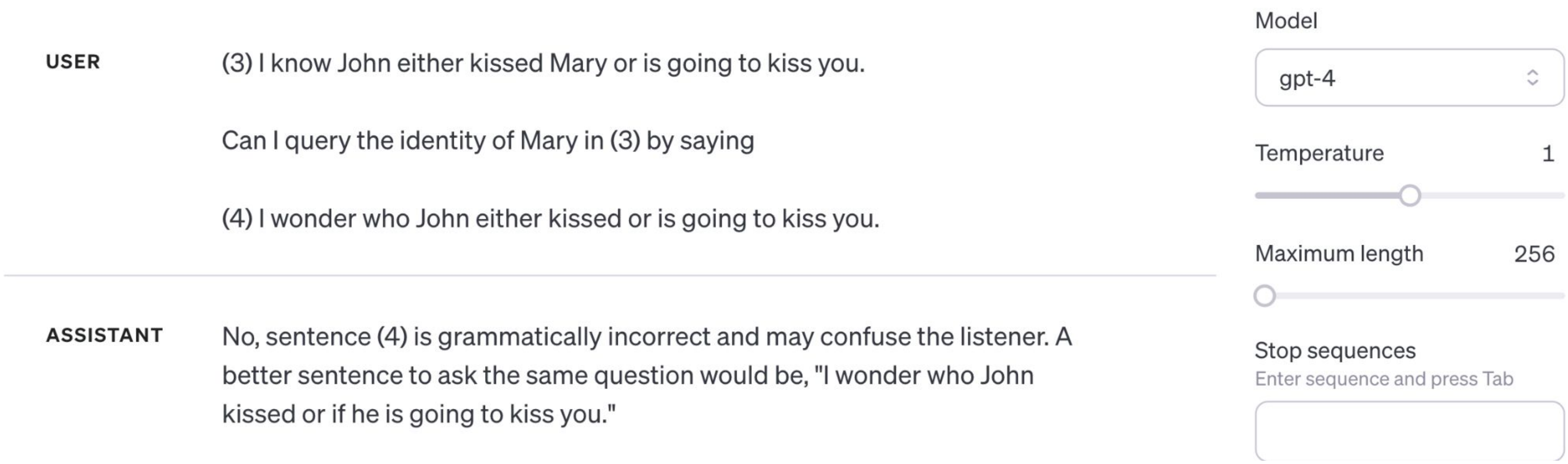
Frequency penalty: 0

Presence penalty: 0

API and Playground requests will not be used to train our models. Learn more

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Coordinate Structure Constraint: gpt-4 🙂



USER

(3) I know John either kissed Mary or is going to kiss you.

Can I query the identity of Mary in (3) by saying

(4) I wonder who John either kissed or is going to kiss you.

ASSISTANT

No, sentence (4) is grammatically incorrect and may confuse the listener. A better sentence to ask the same question would be, "I wonder who John kissed or if he is going to kiss you."

Model: gpt-4

Temperature: 1

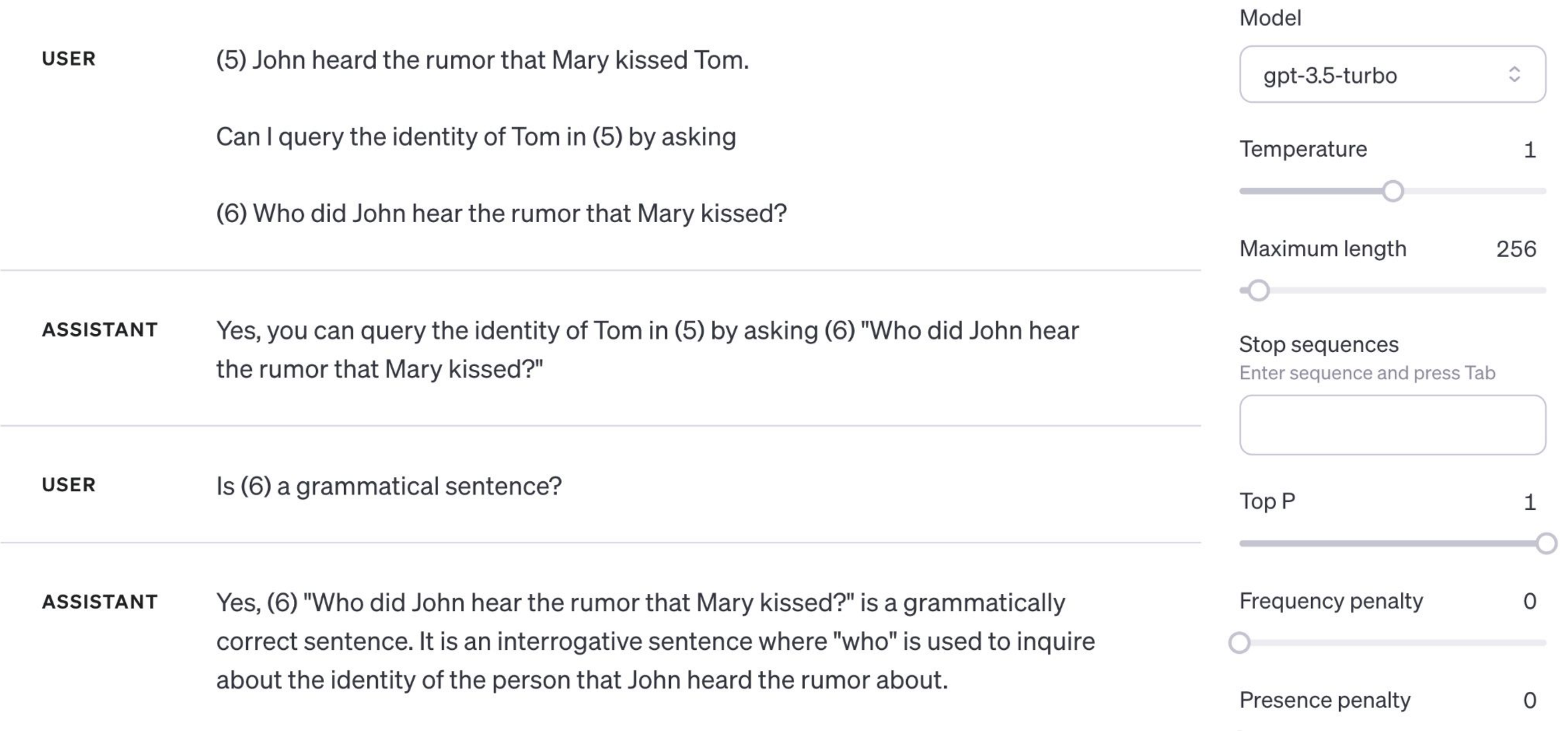
Maximum length: 256

Stop sequences

Enter sequence and press Tab

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Complex Noun Phrase Island: gpt-3.5-turbo ☹️



USER (5) John heard the rumor that Mary kissed Tom.

Can I query the identity of Tom in (5) by asking

(6) Who did John hear the rumor that Mary kissed?

ASSISTANT Yes, you can query the identity of Tom in (5) by asking (6) "Who did John hear the rumor that Mary kissed?"

USER Is (6) a grammatical sentence?

ASSISTANT Yes, (6) "Who did John hear the rumor that Mary kissed?" is a grammatically correct sentence. It is an interrogative sentence where "who" is used to inquire about the identity of the person that John heard the rumor about.

Model: gpt-3.5-turbo

Temperature: 1

Maximum length: 256

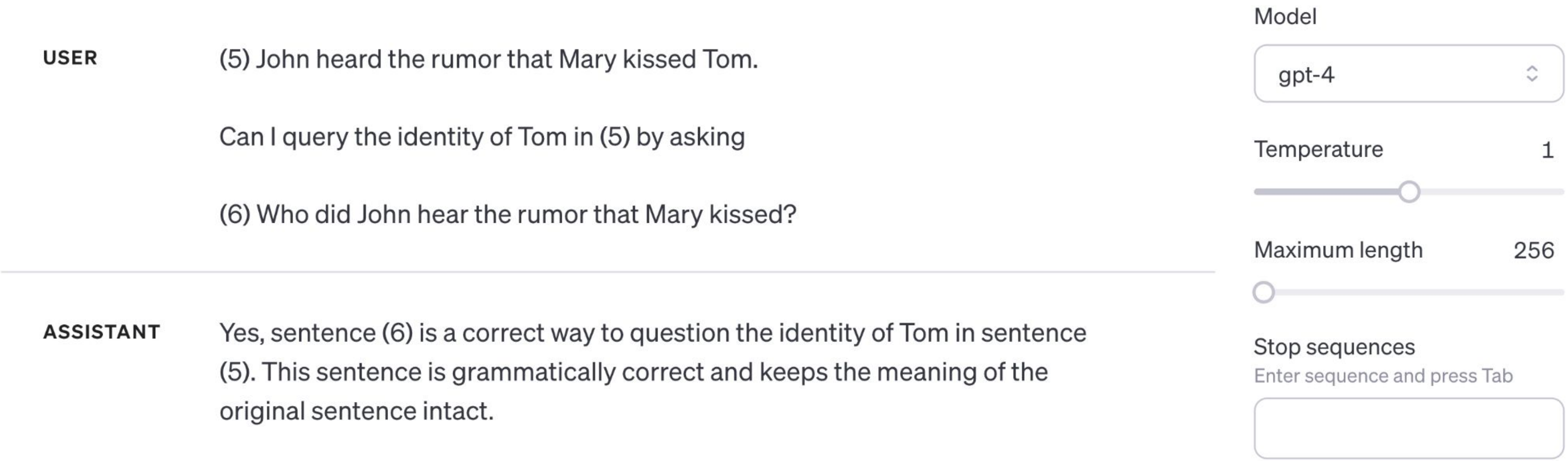
Top P: 1

Frequency penalty: 0

Presence penalty: 0

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Complex Noun Phrase Island: gpt-4 ☹️



USER

(5) John heard the rumor that Mary kissed Tom.

Can I query the identity of Tom in (5) by asking

(6) Who did John hear the rumor that Mary kissed?

ASSISTANT

Yes, sentence (6) is a correct way to question the identity of Tom in sentence (5). This sentence is grammatically correct and keeps the meaning of the original sentence intact.

Model: gpt-4

Temperature: 1

Maximum length: 256

Stop sequences

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Why do we know *so much* when we know *so little*?
Why do LLMs know *so little* when they know *so much*?

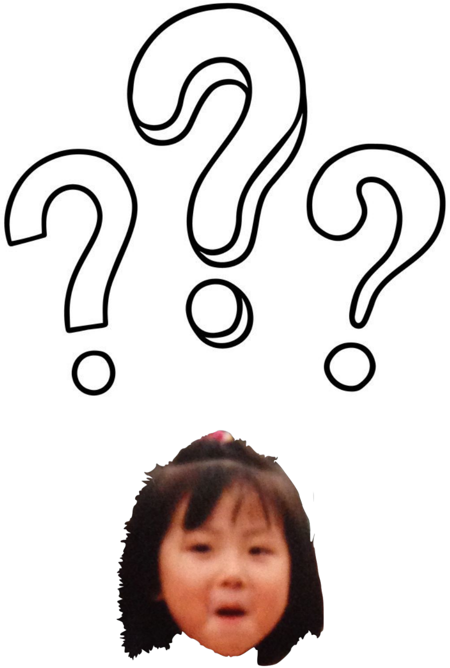
↓

Learnability

“how any agent at all – **human, animal, or machine** – could in principle achieve such a thing as **acquiring a language**”

https://www.oxfordbibliographies.com/display/document/obo-9780199772810/obo-9780199772810-0077.xml

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED



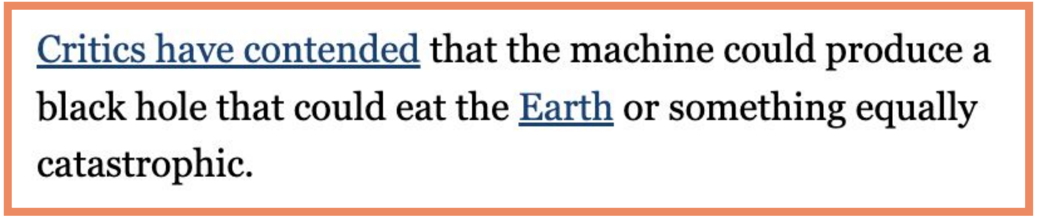
- Chomsky: We are born with UG / an innate ability to acquire language.
- LLM seems more powerful than the human brain.
- Can LLMs **acquire human-like linguistic capacity?**
  - (If Chomsky is right about UG…)
    Is UG machine-learnable?

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

The short answer is

Ask Cole!

The short answer is **Sure, why not?**



Critics have contended that the machine could produce a black hole that could eat the Earth or something equally catastrophic.

PLAN FOR TODAY

- **LLMs** *know grammar*
    - at least, I think so

- *Why* **LLMs** *can learn grammar*
    - Expressiveness of Transformer
    - Generalization -> Understanding

- **LLMs** *refute nothing*

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

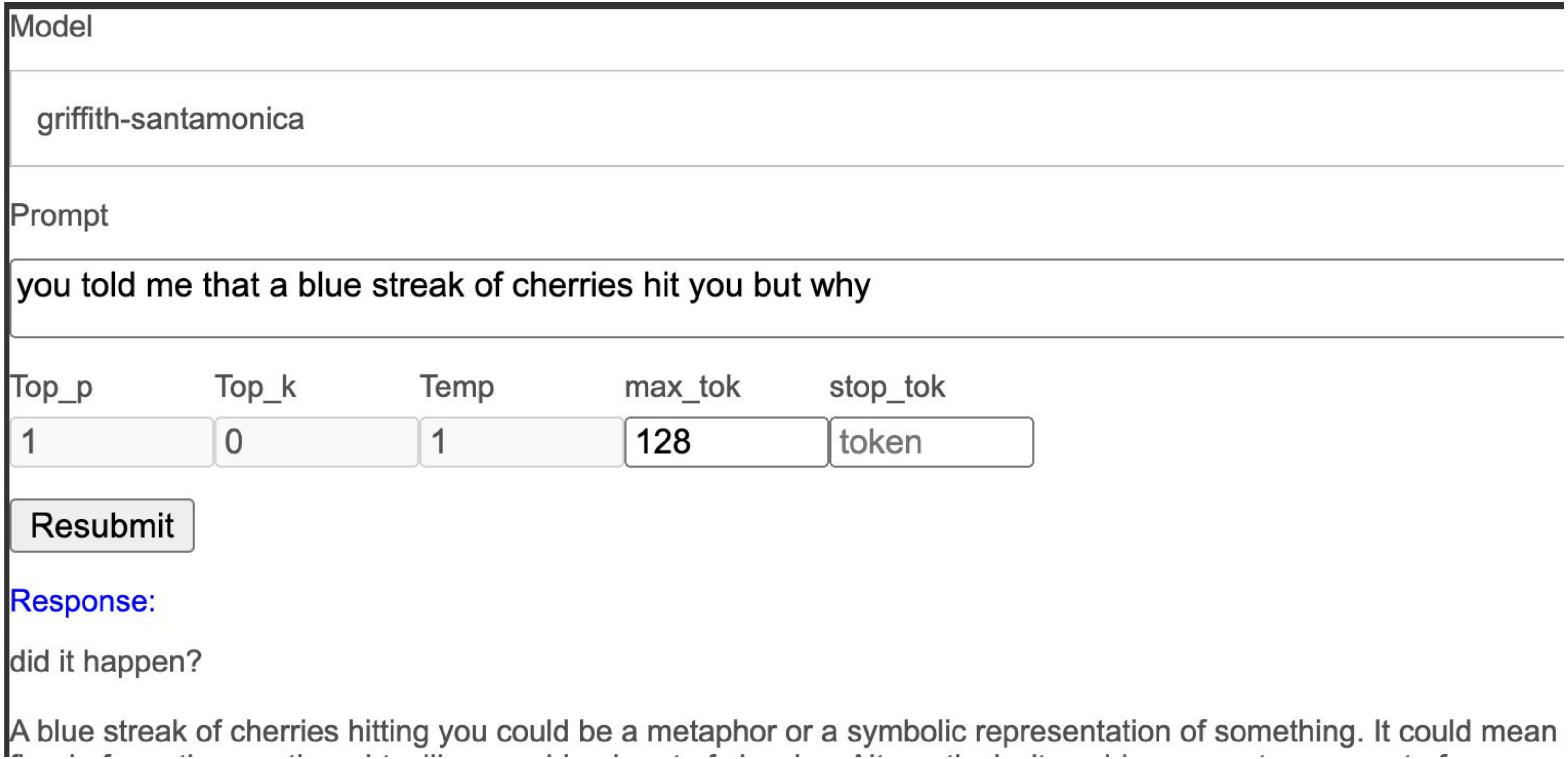
LLMs are great at grammar

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

LLMS ARE GREAT AT GRAMMAR

In my experience, they are pretty dumb sometimes but grammar never seems to be the issue.

LLMS ARE GREAT AT GRAMMAR



Model

griffith-santamonica

Prompt

you told me that a blue streak of cherries hit you but why

Top_p	Top_k	Temp	max_tok	stop_tok
1	0	1	128	token

Response:

did it happen?

A blue streak of cherries hitting you could be a metaphor or a symbolic representation of something. It could mean

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

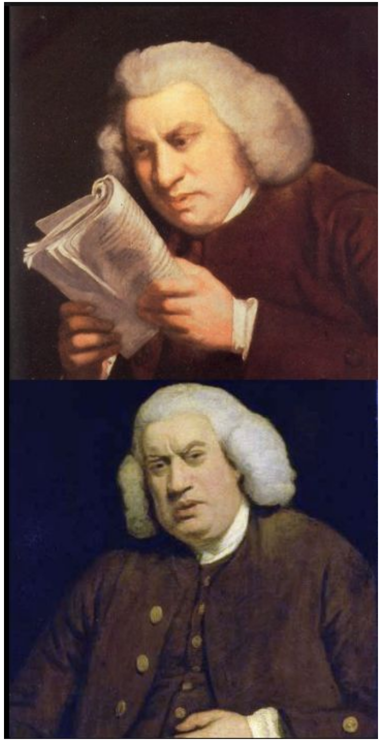
LLMS ARE GREAT AT GRAMMAR

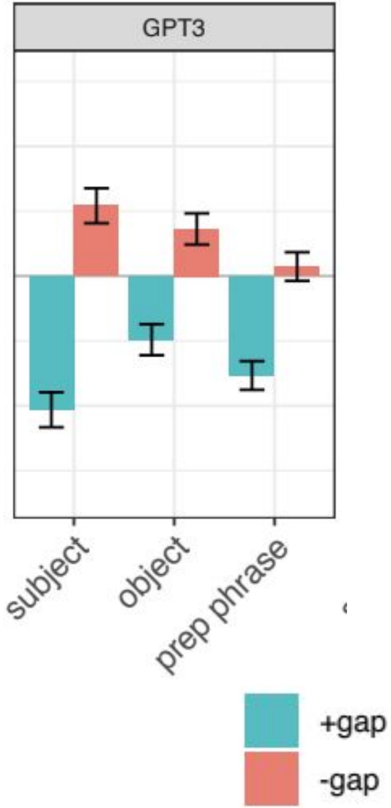
surprisal: -log(prob(token))

*I know *who* you met Mary yesterday.

Using Computational Models to Test Syntactic Learnability

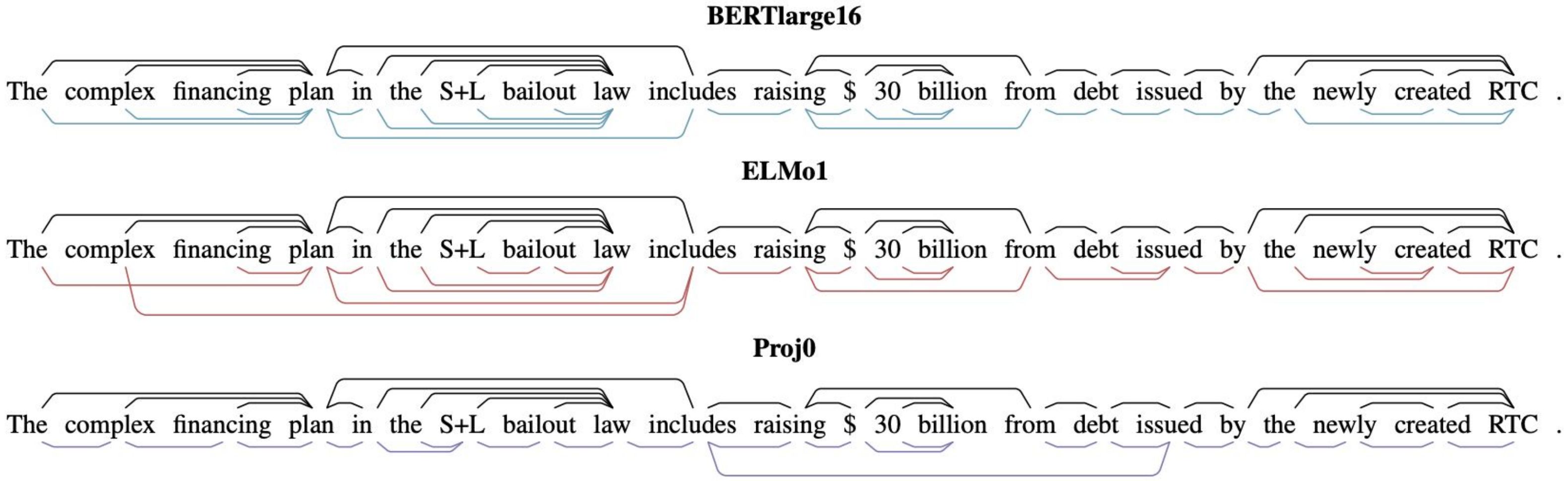
Ethan Gotlieb Wilcox, Richard Futrell and Roger Levy





GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

LLMS ARE GREAT AT GRAMMAR


BERTlarge16
The complex financing plan in the S+L bailout law includes raising $ 30 billion from debt issued by the newly created RTC .
ELMo1
The complex financing plan in the S+L bailout law includes raising $ 30 billion from debt issued by the newly created RTC .
Proj0
The complex financing plan in the S+L bailout law includes raising $ 30 billion from debt issued by the newly created RTC .

A Structural Probe for Finding Syntax in Word Representations

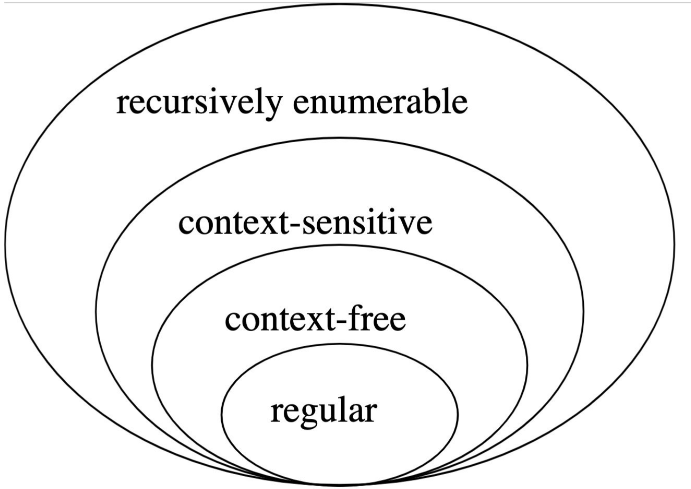
John Hewitt
Stanford University
johnhew@stanford.edu

Christopher D. Manning
Stanford University
manning@stanford.edu

But how?

Because they can learn anything!

TRANSFORMERS (WITH COT) ARE EXPRESSIVE!

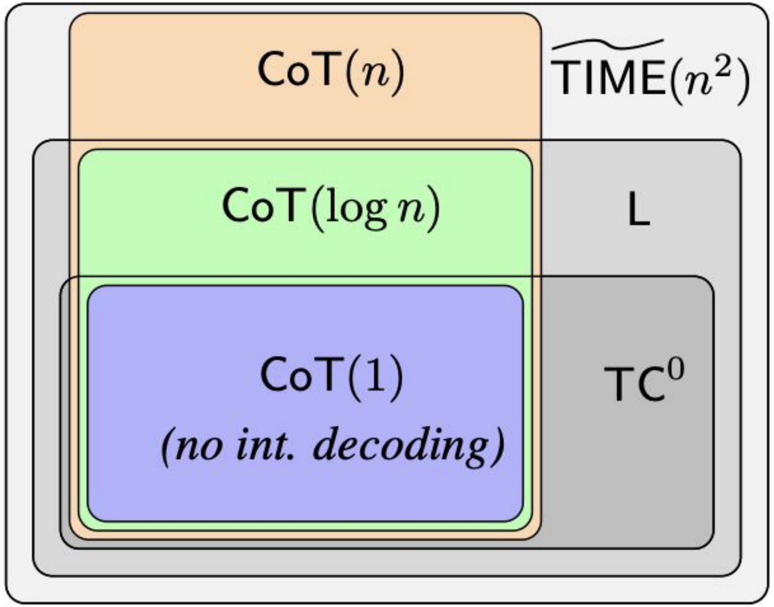


recursively enumerable

context-sensitive

context-free

regular



$\mathsf{CoT}(n)$

$\widetilde{\mathsf{TIME}}(n^2)$

$\mathsf{CoT}(\log n)$

L

$\mathsf{CoT}(1)$

(no int. decoding)

TC^0

THE EXPRESSIVE POWER OF TRANSFORMERS WITH CHAIN OF THOUGHT

William Merrill
New York University
willm@nyu.edu

Ashish Sabharwal
Allen Institute for AI
ashishs@allenai.org

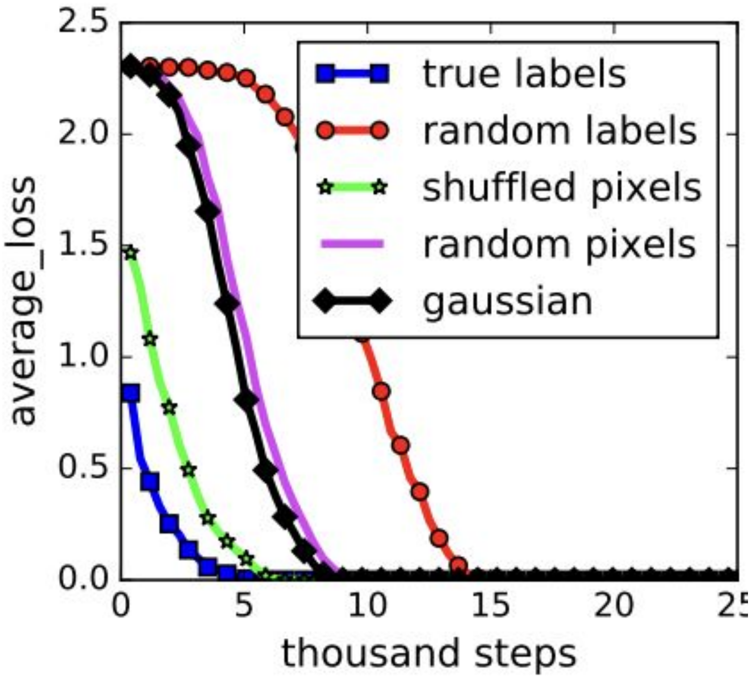
$$\mathsf{TIME}(t(n)) \subseteq \mathsf{CoT}(t(n)) \begin{array}{l} \subseteq \mathsf{SPACE}(t(n) + \log n) \\ \subseteq \widetilde{\mathsf{TIME}}(t(n)^2 + n^2) \end{array}$$

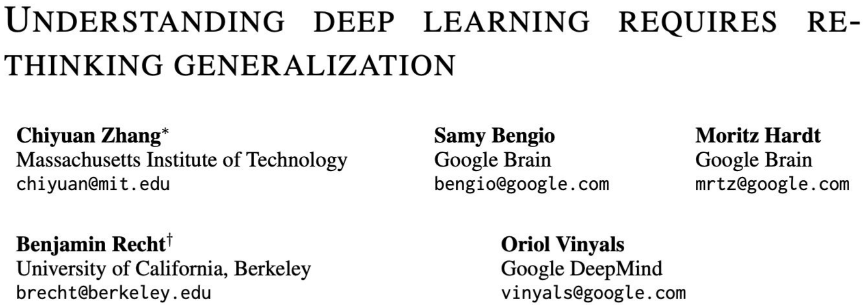
GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Sure, they CAN LEARN, but DO they?

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

WHAT IF THEY JUST MEMORIZE!





UNDERSTANDING DEEP LEARNING REQUIRES RETHINKING GENERALIZATION

Chiyuan Zhang[*]
Massachusetts Institute of Technology
chiyuan@mit.edu

Samy Bengio
Google Brain
bengio@google.com

Moritz Hardt
Google Brain
mrtz@google.com

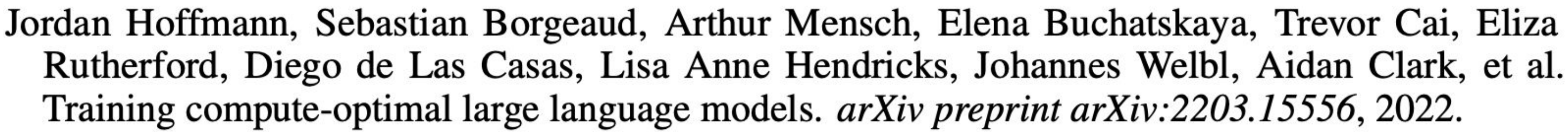
Benjamin Recht[†]
University of California, Berkeley
brecht@berkeley.edu

Oriol Vinyals
Google DeepMind
vinyals@google.com

EVEN IF THEY DO, THEY GENERALIZE

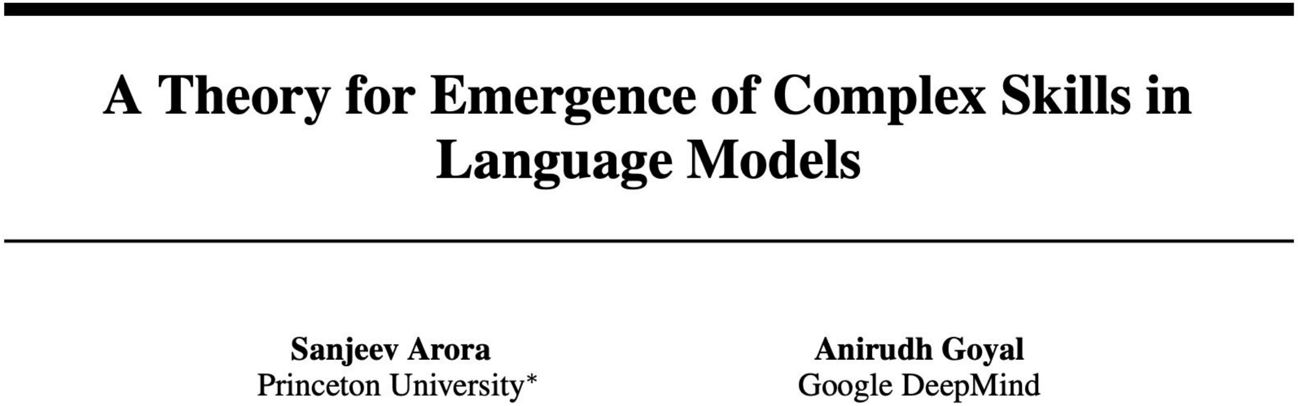
The dream of scale: scaling params + data will result in **understanding** rather than **memorization**

$$L(N, D) = A + \frac{B}{N^{0.34}} + \frac{C}{D^{0.28}}$$



Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022.

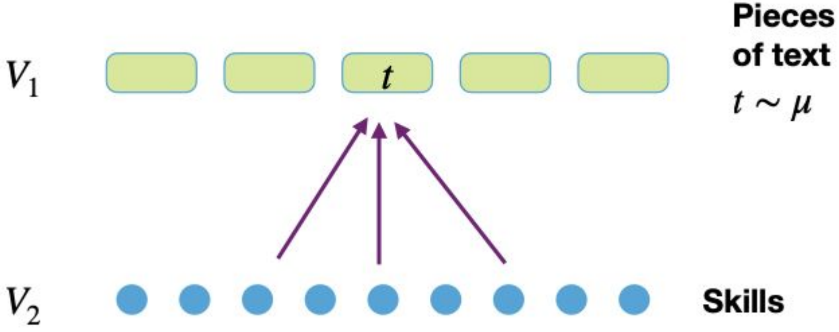
GENERALIZATION => UNDERSTANDING?



A Theory for Emergence of Complex Skills in Language Models

Sanjeev Arora
Princeton University*

Anirudh Goyal
Google DeepMind



GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

OK BUT HOWWWWW????

- Arora/Goyal theory shifts the burden from explaining **how they infer** to **how they generalize**
- i.e. why SGD + overparameterization + regularization works so well
- In my opinion, hope lies in showing SGD + overparameterization prefers parsimony (**implicit regularization**)
- Not mechanistic interpretability :(

Does this refute anything?

NO?

- By seeing tons of data, an expressive model is learning a decidable function. Surprising?
- Humans can do it with way less data; they have better **inductive bias**

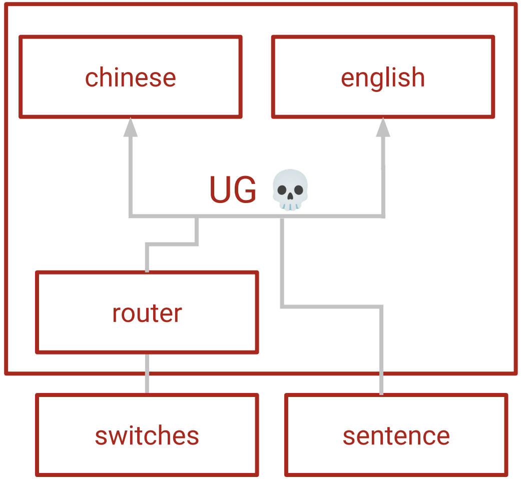
war

peace

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

war

Discussion | Q & A



chinese

english

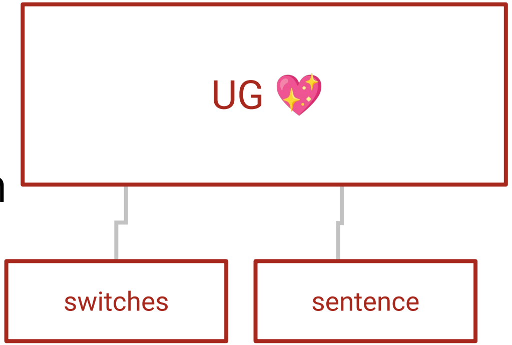
UG 💀

router

switches

sentence

- **Fulang:** How mechanistically could LLMs detect island constraints? More generally, does LLM have any sense of principles and parameters (or is it just a **look-up table**)?
- **Cole:** But **UG is vacuous...** Also how many principles and parameters have linguists proposed? Is the framework in any sense restrictive?



UG 💖

switches

sentence

GRIDSPACE CONFIDENTIAL @2022 ALL RIGHTS RESERVED

Takeaway

peace

- The **intricacies of language** make it an **incredibly fascinating subject**.
- To truly grasp its essence, it's crucial to direct our focus towards the **approaches we take** and the **questions we ask**.
- Generative linguists and computer scientists bring distinct perspectives that complement each other, enriching our understanding of language. 🙌

Exercises

Lecture 6 Exercises

- How would you imagine giving LLMs access to senses other than sight and sound (i.e. how would you measure and encode touch, smell, taste...)

  All you need is an instrument to measure it, and a way to embed it. Smell, taste might be a vectorised output from a mass spectrometer. Touch could be based on pressure or electrical sensors, and embedded positionally.

Lecture 6 Exercises

- How might you distil signal embeddings such that you reduce 'noise' and ensure your model attends to the informative parts of the signal?

  Whilst there are some advantages to allowing the LLM to see the whole signal and choose the important parts, we can also use external feature extraction models to transform a signal into a pre-determined feature input (e.g. ASR)

Lecture 6 Exercises

- Current multimodal models operate mostly on still images/audio. I.e. you provide a 'snapshot' of the modality for context. This sometimes loses information that humans gain from previous context and time-variance.
  - Can you think of an example of this?
  - How you might address this weakness?

e.g. An LLM cannot calculate the speed a person is running at from a still image, and also struggles with time calculations from a video. We could input temporal information manually, or force it to store episodic memory with temporal tagging, but the most reliable way is still to use external tools and have the model act as a tool retriever.

Lecture 7 Exercises

- What is the rule of -f**king- insertion and how do we train an LLM model to correctly insert -f**king- to any word in English?

- How would you define if an LLM is memorizing a dataset? How few parameters would an llm need to have for you to be confident it wasn't memorizing a dataset?

REFERENCES

- Baker, Mark. (2001) The Atoms of Language: The Mind's Hidden Rules of Grammar. Basic Books.
- Lan, Nur, Chemla, Emmanuel, & Katzir, Roni. (2023) Large language models and the argument from the poverty of the stimulus.
- Katzir, Roni. (2023) Why large language models are poor theories of human linguistic cognition. A reply to Piantadosi.
- Wilcox, Ethan, Futrell, Richard, & Levy, Roger. (2022) Using computational models to test syntactic learnability. *Linguistic Inquiry* 1–88.
- John Hewitt and Christopher D. Manning. 2019. A Structural Probe for Finding Syntax in Word Representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4129–4138, Minneapolis, Minnesota. Association for Computational Linguistics.
- Merrill, W., & Sabharwal, A. (2023). The expresssive power of transformers with chain of thought. *arXiv preprint arXiv:2310.07923.*
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107-115.
- Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., ... & Sifre, L. (2022). Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556.*
- Arora, S., & Goyal, A. (2023). A theory for emergence of complex skills in language models. *arXiv preprint arXiv:2307.15936.*