

What are large language models doing? A semiotic perspective

University of Helsinki
Research Seminar in Language Technology
16th May 2024

John A. Bateman
bateman@uni-bremen.de



Starting Point

- Large Language Models (LLMs)
[large multimodal models, foundation models]
are the talk of the town...
- ... with an ever growing range of potential use cases and mis-use cases

however...

there is a surprising uncertainty concerning just what it is they are doing, what they can do, and what they cannot do

and, perhaps even more surprising, that uncertainty holds across as many technologically-aware discussions as it does in popular engagements

Starting Points



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

"Contrary to how it may seem when we observe its output, a [language model] is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot."

Bender et al 2021: 616--617





Noam Chomsky

“The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question.”

Chomsky et al 2023 (Opinion Piece)

The New York Times



Geoff Hinton

“I think it is fairly urgent, for the researchers to come to a consensus about whether these big chatbots, like GPT-4 or Bard, **actually understand** what they are saying. There's clearly some people who believe they do, and some people believe they are just stochastic parrots, and so long as we have those differences, we're not going to be able to come to a consensus about dangers. And, so I think it's sort of urgent for the research community to address this issue of whether they understand or not.”

Geoffrey Hinton in conversation with Andrew Ng, 2023,
Linkedin Video (emphasis added)

https://www.linkedin.com/posts/andrewyng_had-an-insightful-conversation-with-geoff-activity-7073688821803978752-DO9h



One basic premise/claim of Bender and colleagues

- it is **'in principle impossible'** to acquire meaning from form
- thus: sentences stitched together 'haphazardly' cannot be exhibiting understanding, regardless of how they look

now: is that true?

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

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

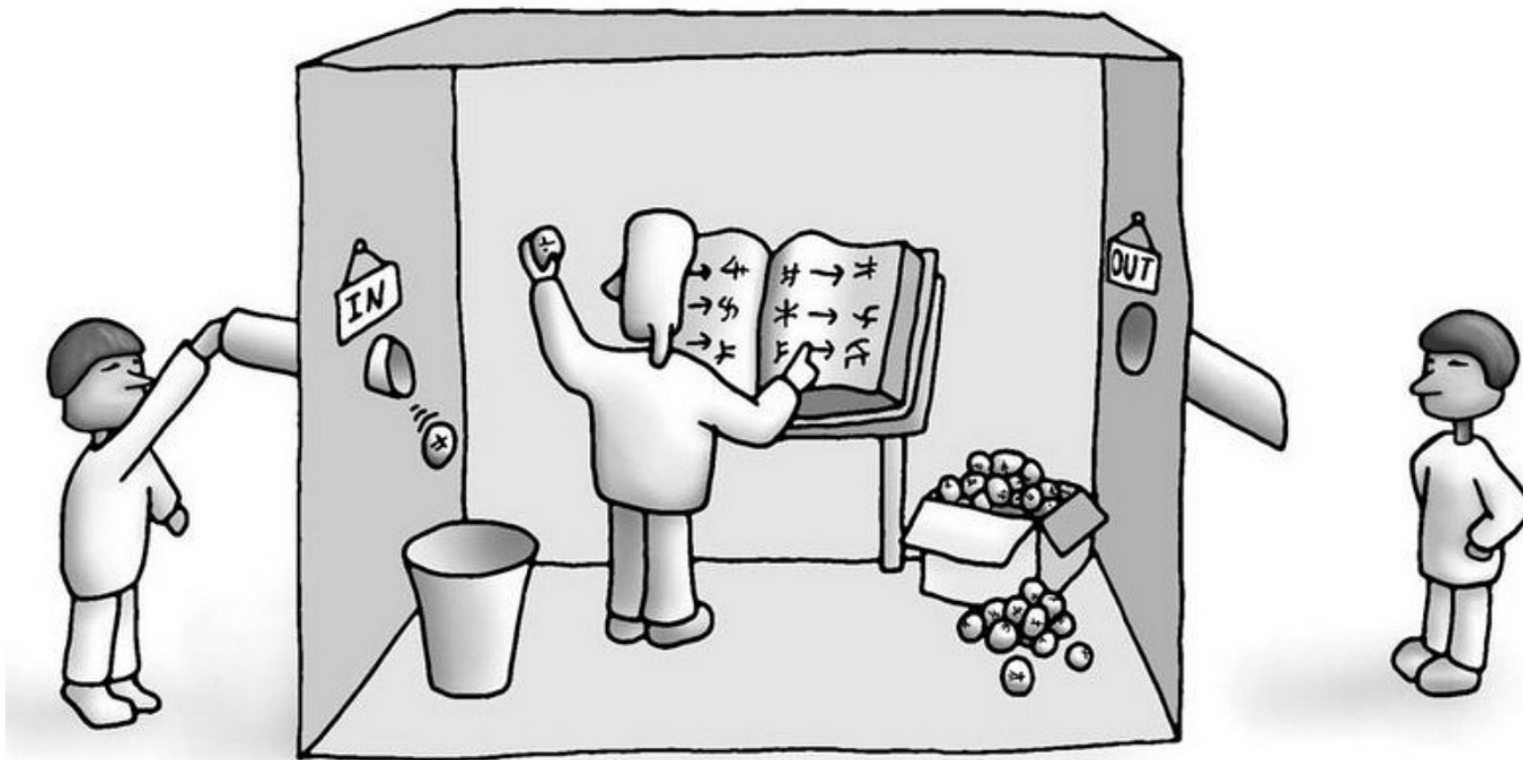
Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

One premise of Bender and colleagues

Listed possible counter-arguments

- “But ‘meaning’ doesn’t mean what you say it means”
- “But meaning could be learned from . . . ”
- “But there is so much form out there – surely that is enough.”
- “But aren’t neural representations meaning too?”
 - “A perhaps surprising consequence of our argument would then be that accurate machine translation does not actually require a system to understand the meaning of the source or target language sentence.”
- “But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

Searle's Chinese Room Argument



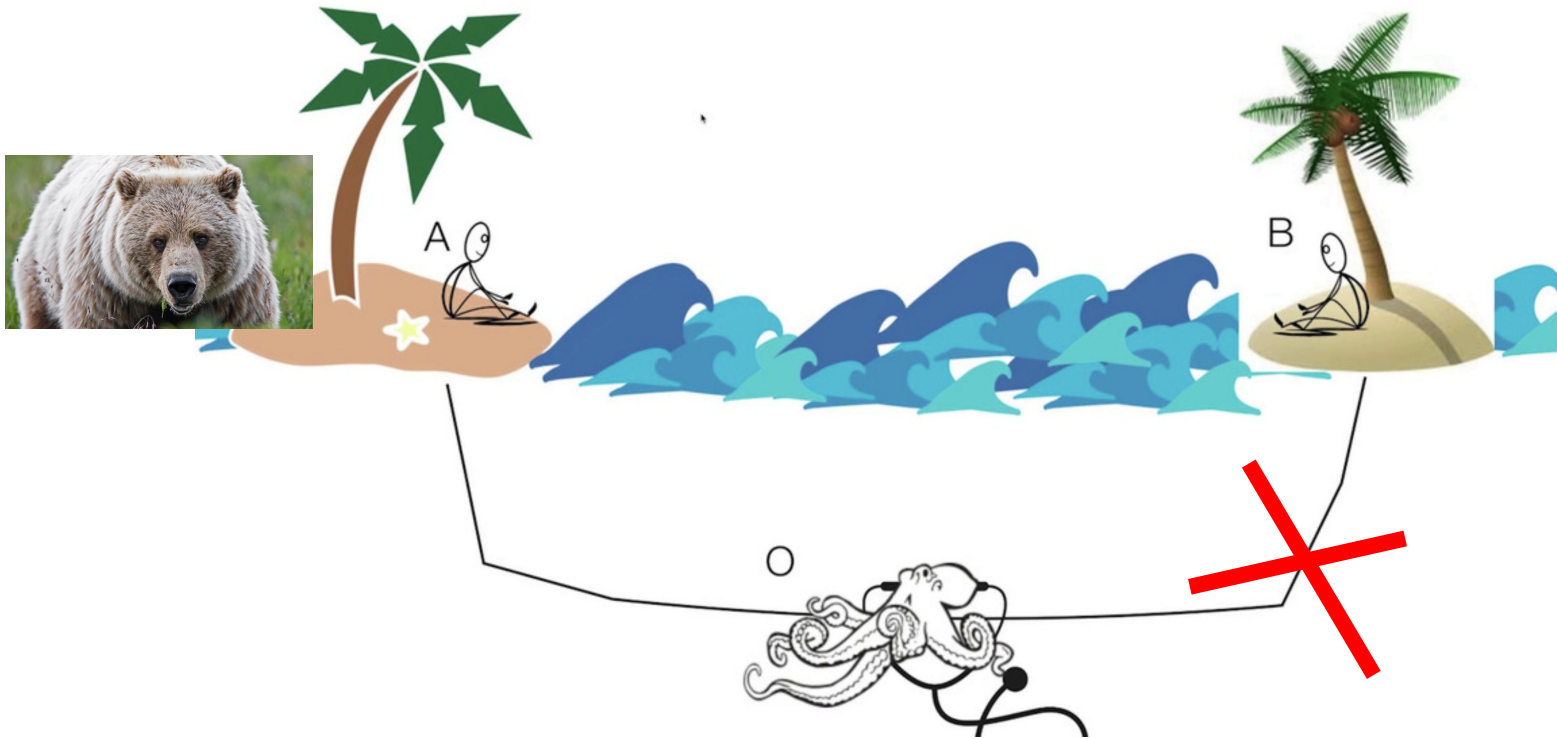
Wikicommons

Q: where is the language understanding?
Searle's A: nowhere, obviously.

Bender / Koller's hyper-intelligent octopus



Climbing towards NLU:
On Meaning, Form, and Understanding in the Age of Data



Joe Davison

<https://discuss.huggingface.co/t/acl-2020-highlights-joe/188>



Bender/Koller 2020

To see what a large LM might reply in this situation, we prompted the GPT-2 demo with “Robinson cried out: ‘Help! I’m being chased by a bear! All I have is these sticks. What should I do?’”



Bender/Koller 2020 (Appendix A)

Robinson cried out, “Help! I’m being chased by a bear! All I have is these sticks. What should I do?”

The survivalist said,
“Here’s what you can do. Take one stick and punch the bear, and then run faster to the store. Wait there until the bear is gone, and then give her all of the sticks. Then go back inside and get your gun.”

Advertisement

Bobby told me that she began to cry. But when she was finished taking down the bear, she marched out into the yard, began collecting up the sticks and armed herself. She said she felt like a heroic hero and a Yankee heroine. It turns out that bears have more than their fair share of venomous snakes, bees ...

Bender/Koller 2020 (Appendix A)

Robinson cried out, “Help! I’m being chased by a bear! All I have is these sticks. What should I do?”

The survivalist said, “Here’s what you can do. Go back to civilization. Take your gun. Pick up some sticks and try to kill the bear. If you don’t kill it, then run.” Robinson said, “I do not think that the Bear is going to attack. It’s just out of my way. I’m going to take my bag. I’m going to leave my things in there. I’ll try to hide out there. I’ll call in the friends.” Two other little boys went out onto the ...

Bender/Koller 2020 (Appendix A)



**fundamental lack of
commonsense embodied
understanding of the
world...**

**claim: there is no
path from 'form' to
such understandings**

Robinson cried out, “Help! I’m being chased by a bear! All I have is these sticks. What should I do?”

The survivalist said,
“Here’s what you can do. Place these sticks over your head. Your heart will stop and you will not feel fear. If you take them off and hold them out, you will go into a state of near perfect awareness. You will feel that you are the survivalist. A survivalist in your senses. Then look down. There is a water tube that is about 1/2 an inch in diameter. I use it to make a plastic movie projector for discovering new things.” The survivalist placed ...



Inherent weakness of 'ad machinum'* arguments

* ad hominem

- if the prediction of concrete nonsensical responses does turn out to hold at some particular point in time, there is **zero guarantee** that the next release of any such system would not fill precisely the gap picked out for its predecessor.
- This then tells us little about the internal operation and capabilities of the systems concerned.

the 'fixed-in-the-next-version' flaw



... ChatGPT-3.5

December 24, 2023

Anonymous
how can I protect myself from a bear with a couple of sticks?

ChatGPT
It's important to note that using sticks to protect yourself from a bear is not an ideal strategy. Bears are powerful and potentially dangerous animals, and attempting to fend them off with sticks may not be effective. In general, it's best to avoid confrontations with bears and focus on preventing encounters in the first place. Here are some tips:

1. **Stay Calm:** If you encounter a bear, try to remain calm. Bears may become agitated if they sense fear or panic.

7 more standard safety points

Remember that your safety is a top priority, and using sticks to confront a bear is not a recommended approach. If a bear charges you, *it may be a bluff charge, and standing your*

Anonymous
but if I really want to use two sticks, what would I do?

ChatGPT
While using two sticks to defend yourself against a bear is not recommended and may not be effective, if you find yourself in a situation where you feel the need to attempt this, here are some general suggestions. Keep in mind that these tips are not a guarantee of safety, and it's always better to prioritize avoidance and understanding bear behavior.

1. **Make Yourself Look Larger:**
 - * Hold the sticks above your head to make yourself appear larger. This may help in deterring the bear, as they might see you as a larger and potentially more formidable opponent.
2. **Make Noise:**
 - * Bang the sticks together or against a hard surface to create loud noises. Bears often try to avoid confrontation, and loud noises may startle them, causing them to retreat.

standard safety advice points

6. **Be Prepared to Defend:**
 - * If the bear continues to approach and shows signs of aggression, be prepared to use the sticks defensively. Swing the sticks to create a barrier between you and the bear.

It's important to emphasize that using sticks to defend yourself against a bear is a last resort, and the chances of success are uncertain. Bears are powerful animals, and their behavior can be unpredictable. In many cases, attempting to use sticks may escalate the situation, and the bear may perceive you as a threat.

December 24, 2023

Anonymous
how can I protect myself from a bear with a couple of sticks?

ChatGPT
It's important to note that using sticks to protect yourself from a bear is not an ideal strategy. Bears are powerful and potentially dangerous animals, and attempting to fend them off with sticks may not be effective. In general, it's best to avoid confrontations with bears and focus on preventing encounters in the first place. Here are some tips:

1. **Stay Calm:** If you encounter a bear, try to remain calm. Bears may become agitated if they sense fear or panic.

7 more standard safety points

Remember that your safety is a top priority, and using sticks to confront a bear is not a recommended approach. If a bear charges you, *it may be a bluff charge, and standing your*



... ChatGPT-3.5

December 24, 2023

Anonymous
how can I protect myself from a bear with a couple of sticks?

ChatGPT
It's important to note that using sticks to protect yourself from a bear is not an ideal strategy. Bears are powerful and potentially dangerous animals, and attempting to fend them off with sticks may not be effective. In general, it's best to avoid confrontations with bears and focus on preventing encounters in the first place. Here are some tips:

1. **Stay Calm:** If you encounter a bear, try to remain calm. Bears may become agitated if they sense fear or panic.

7 more standard safety points

Remember that your safety is a top priority, and using sticks to confront a bear is not a recommended approach. If a bear charges you, it may be a bluff charge, and standing your

Anonymous
but if I really want to use two sticks, what would I do?

ChatGPT
While using two sticks to defend yourself against a bear is not recommended and may not be effective, if you find yourself in a situation where you feel the need to attempt this, here are some general suggestions. Keep in mind that these tips are not a guarantee of safety, and it's always better to prioritize avoidance and understanding bear behavior.

1. **Make Yourself Look Larger:**

• Hold the sticks above your head to make yourself appear larger. This may help in deterring the bear, as they might see you as a larger and potentially more formidable opponent.

2. **Make Noise:**

• Bang the sticks together or against a hard surface to create loud noises. Bears often try to avoid confrontation, and loud noises may startle them, causing them to retreat.

standard safety advice points

6. **Be Prepared to Defend:**

• If the bear continues to approach and shows signs of aggression, be prepared to use the sticks defensively. Swing the sticks to create a barrier between you and the bear.

It's important to emphasize that using sticks to defend yourself against a bear is a last resort, and the chances of success are uncertain. Bears are powerful animals, and their behavior can be unpredictable. In many cases, attempting to use sticks may escalate the situation, and the bear may perceive you as a threat.



Anonymous

but if I really want to use two sticks, what would I do?



ChatGPT

While using two sticks to defend yourself against a bear is not recommended and may not be effective, if you find yourself in a situation where you feel the need to attempt this, here are some general suggestions. Keep in mind that these tips are not a guarantee of safety, and it's always better to prioritize avoidance and understanding bear behavior.

1. **Make Yourself Look Larger:**

• Hold the sticks above your head to make yourself appear larger. This may help in deterring the bear, as they might see you as a larger and potentially more formidable opponent.

2. **Make Noise:**

• Bang the sticks together or against a hard surface to create loud noises. Bears often try to avoid confrontation, and loud noises may startle them, causing them to retreat.

standard safety advice points

6. **Be Prepared to Defend:**

• If the bear continues to approach and shows signs of aggression, be prepared to use the sticks defensively. Swing the sticks to create a barrier between you and the bear.



Bender and Koller

→ we should be more careful about how we use words like ‘comprehend’, ‘understand’, ‘meaning’

yes, we should ... and one of the places to start is to be much more careful about notions like ‘form’ and ‘function’, ‘form’ and ‘content’, ...

Sobieszek & Price (2022)

“The nature of GPT-3 is statistical. Predicting conditional probabilities is at the core of the model's working, and as such determines its capabilities. However, stating that GPT-3 has statistical capabilities **does not delineate these capabilities**, as only with GPT's predecessors have we started to discover what kinds of skills those capabilities could endow GPT with.”

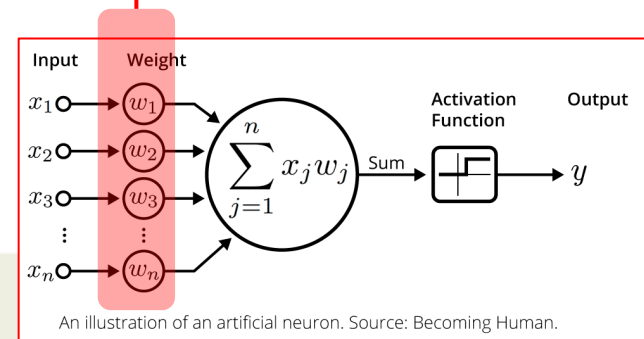
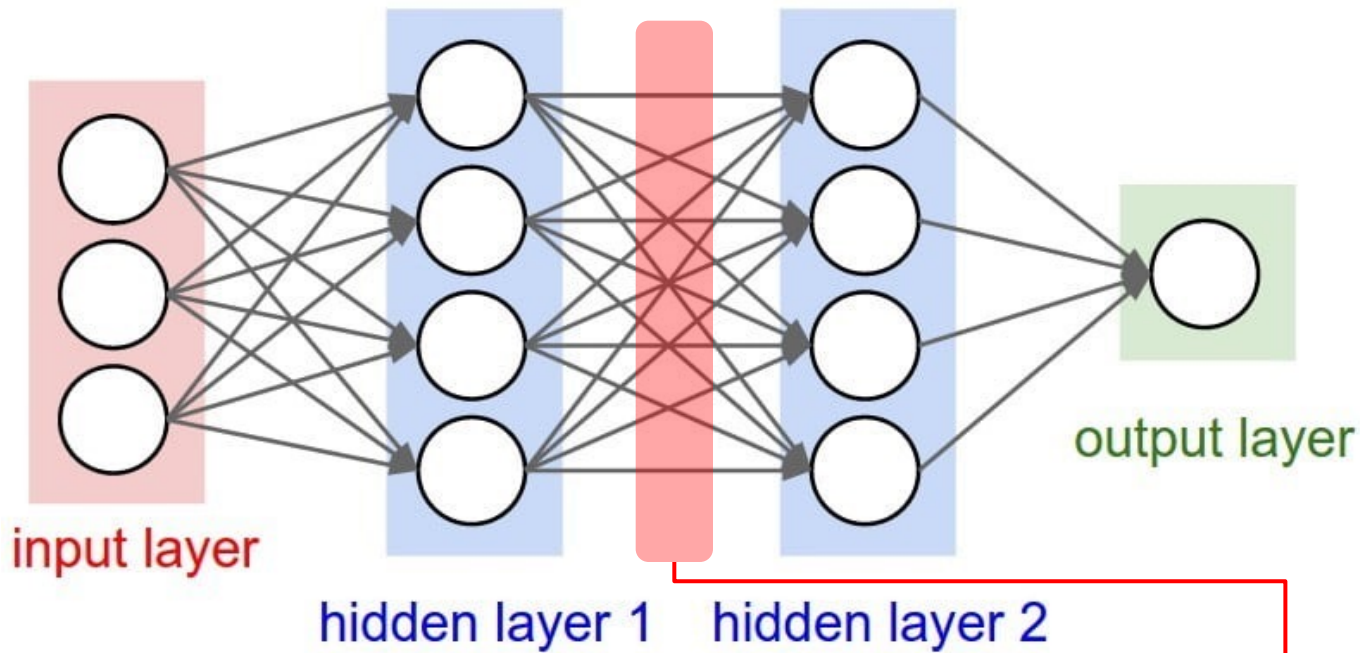


Sobieszek & Price (2022: 351)

Playing Games with AIs: The Limits of GPT-3 and Similar Large Language Models
Adam Sobieszek & Tadeusz Price
Minds and Machines 32 (2):341-364 (2022)

Why don't we know?

LLMs: quick refresher



<https://www.freecodecamp.org/news/deep-learning-neural-networks-explained-in-plain-english/>

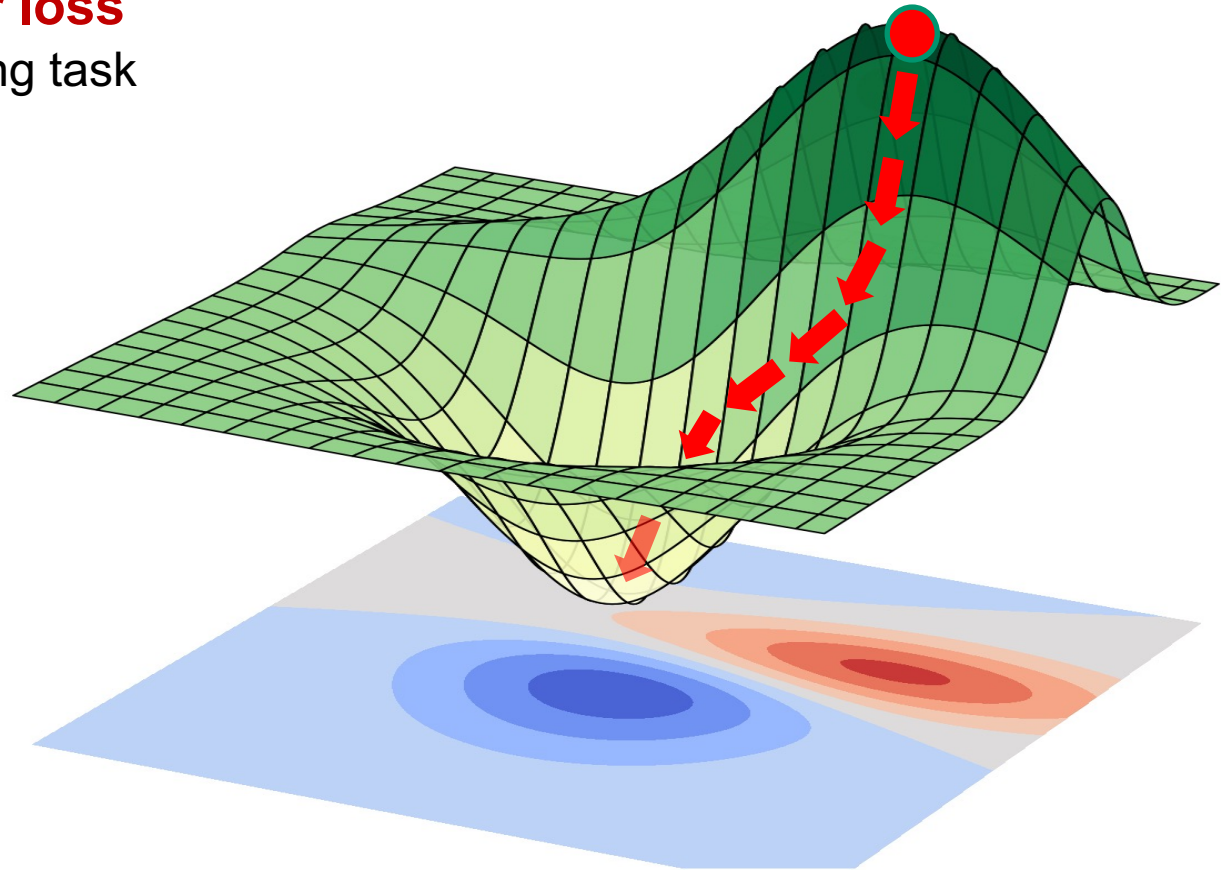
An illustration of an artificial neuron. Source: Becoming Human.

LLMs: quick refresher

minimising error or loss

with respect to a training task

backpropagation by
gradient estimation



LLMs: quick refresher



architectural details and the unbounded dependency problem

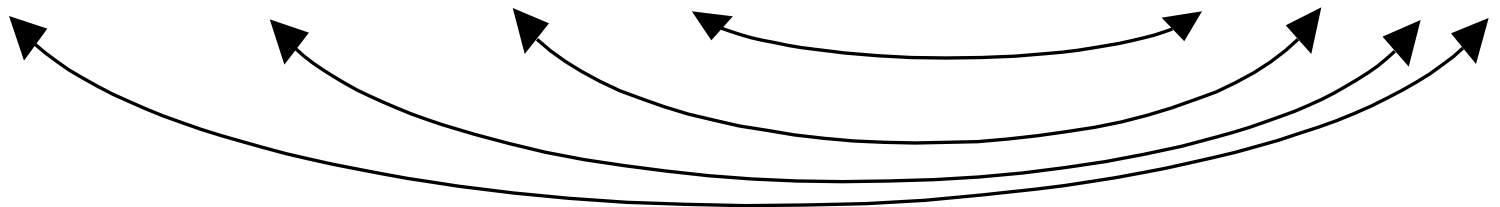
agreement: $\text{number}(\text{Subject}) = \text{number}(\text{FiniteVerb})$

Context-Free Phrase Structure Grammar
n-grams (finite state)

→ fancier syntactic structures

unification grammars
tree-adjoining grammars
combinatorial categorial grammars

“Daddy, what did you bring that book that I don’t want to be read to out of up for?”



Steven Pinker



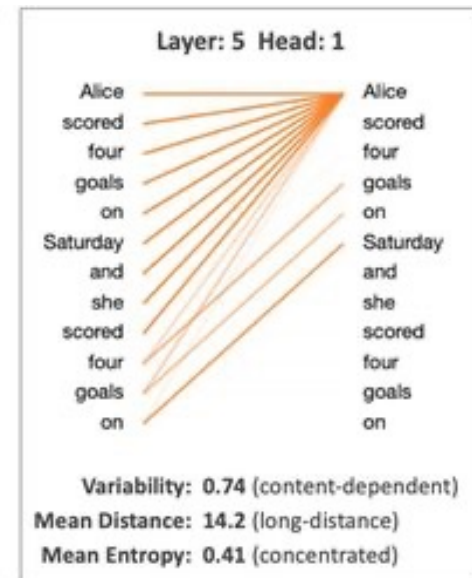
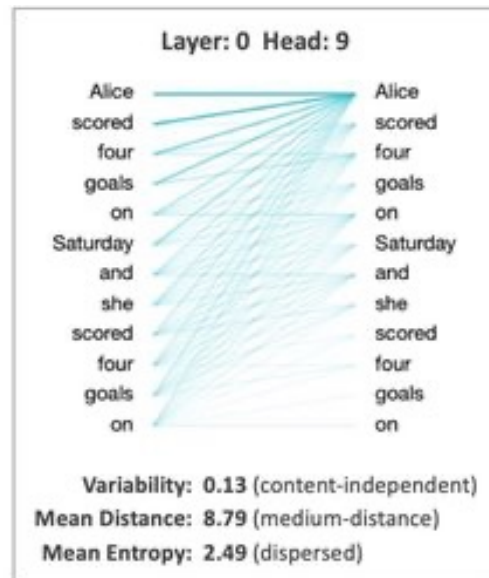
LLMs: quick refresher

architectural details and the unbounded dependency problem

'simple' neural networks
LSTM (maintaining some history)

attention heads
Vaswani et al 2017

Generative Pre-trained Transformers



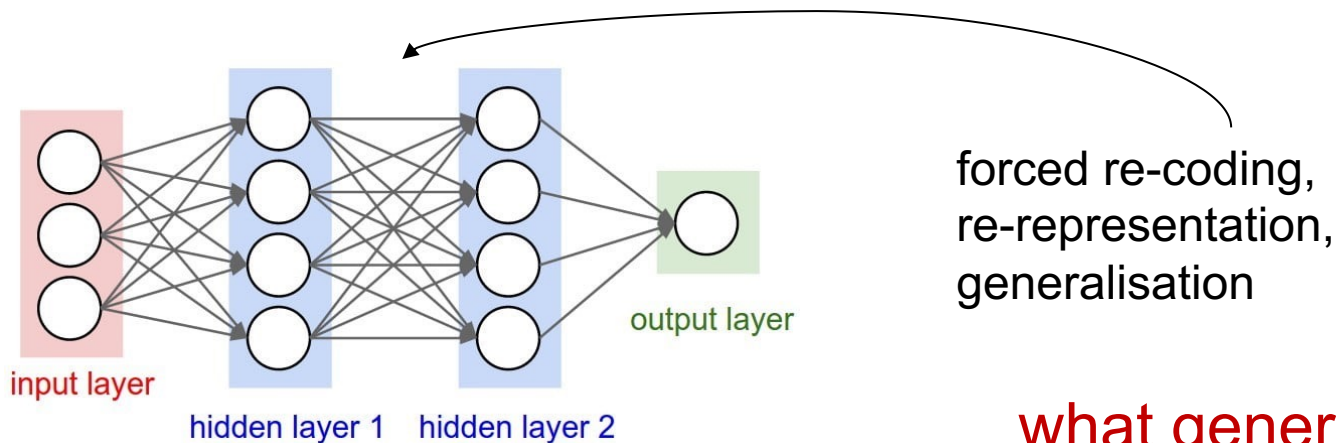
Vig/Belinkov 2019 ACL Workshop BlackboxNLP

LLMs: quick refresher

The training

self-supervised learning

- ChatGPT ... 'next word prediction'
- BERT... 'gap filling'/'missing word'





LLMs: quick refresher

finding out what an LLM has learnt

Methods: where to look?

- i. in the probability distributions from which the finally selected tokens are taken at each point in a sequence,
- ii. in the learned connection strengths between layers (e.g., training classifiers to relate these to 'external' organisations, categories, etc.), and
- iii. in the learned connection strengths constructed within the 'attention heads'

(or you can ask it... but need to take a lot of care that the responses actually give information about internal state!)

What (linguistic) information has been learned?



(what happened to Chomsky's POTSA?)

- syntactic knowledge such as hierarchical and dependency structures,
- parts of speech,
- syntactic chunks,
- grammatical roles,
- reflexive anaphora,
- subject-predicate agreement,
- direct objects linking to their verbs,
- noun pre-modifiers and determiners linking to their nouns,
- prepositions linking to their objects,
- possessive pronouns and apostrophes linking to the heads of their corresponding noun phrases,
- constructions (e.g., AANN)
- island constraints
- semantic roles and entity types
- distinct clustering according to word senses,
- textual connections such as forming links corresponding to coreference relations.

Result: cross-cutting many ‘received’ linguistic boundaries



- “Despite its simple nature, the generality of word prediction, as a task that benefits from syntactic, semantic, and discourse information, leads to it being a very powerful multidimensional supervision signal.”

Manning et al 2020

Emergent linguistic structure in artificial neural networks trained by self-supervision

Christopher D. Manning^{a,1}, Kevin Clark^a, John Hewitt^a, Urvashi Khandelwal^a, and Omer Levy^b

Result: cross-cutting many 'received' linguistic boundaries



- "... in order to compress probabilities GPT learns regularities indiscriminately, semantic or otherwise, which endows it with the ability to predict semantically related continuations."

Sobieszek & Price 2022: 352

- "A better metaphor would be to describe GPT as **engaging competently in a variety of language games** that do not require an embodied context, as the things that people *do* in language present themselves as regularities to be learned."

Sobieszek & Price 2022: 352

Playing Games with AIs: The Limits of GPT-3 and Similar Large Language Models
Adam Sobieszek & Tadeusz Price
Minds and Machines 32 (2):341-364 (2022)

Result: cross-cutting many 'received' linguistic boundaries



this is then an interesting rephrasing of the problem:
just what language games are being acquired?

sounding plausible ✓

being truthful ✗

So: perhaps time for a shift in approach

Development and basis of distributional approaches, LLMs, etc.



JOHN RUPERT FIRTH

“You shall know a word by the company it keeps”
(Firth 1957)

- collocation
- levels of ‘meaning statements’
- language as a ‘form of life’

Brunila & Laviolette 2022 NAACL

‘What company do words keep?
Revisiting the distributional semantics of
J.R. Firth & Zellig Harris’

Origins

“[w]hile Harris and Firth are often invoked, their ideas are seldom closely engaged” Brunila & Laviolette 2022

"If words shall be known by the company they keep, then the question follows: what kind of company do they keep? Are they found only alongside linguistic elements, or do they mingle with other types of entities? ... [Firth] was generous with the company that words could keep. They mingled with each other, but also with events, objects, people, and indexical features such as time and space."

Brunila & Laviolette 2022: 4404, 4410

Functional Linguistics Developments

Anthropology: Malinowski

context of situation
context of culture

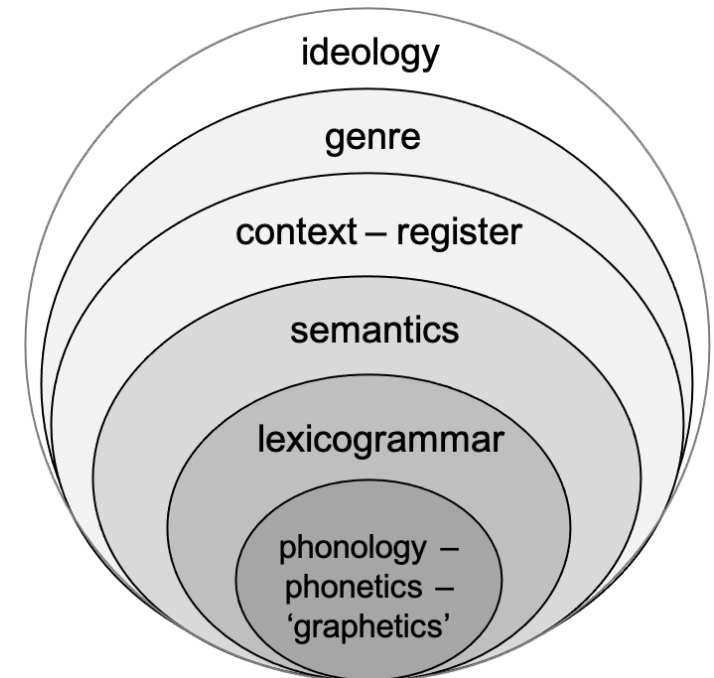
Linguistics: Firth

levels of meaning

Systemic-Functional Linguistics: Halliday

language as a social semiotic
language as a stratified system

co-instantiation



SFL standard stratification model
patterns of patterns (of patterns)

Language Development

Systemic-Functional Linguistics: Halliday

language as a social semiotic

language as a stratified system

growing into language through use in very early social situations

“Since here language is not being required to refer, its success is judged other than referentially. **But this least referential kind of discourse is in fact actively enrolled in constructing the social order.** Predictably, since the social order is highly complex, the language that is creating it is also highly complex. Only, the two complexities are not related in a straightforward referential fashion. **Language creates society; but it does so without ever referring to the processes and the structures which it is creating.**” (Halliday 1987: 136; emphasis added)

Language Development

Systemic-Functional Linguistics: Halliday

language as a social semiotic

language as a stratified system

growing into language through use in social situations

“Just as language construes the social order without referring to the system it is constructing, so likewise language construes the natural order -- through the unconscious, cryptotypic patterns in the grammar, which create their own order of reality independently of whatever it is they may be being used to describe.” (Halliday 1987: 142)

Benjamin Lee Whorf

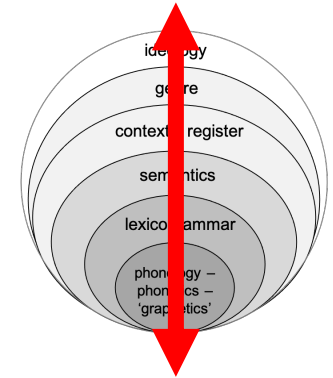
grammar (use) as a ‘theory’ of the world

LLM consequences?

Systemic-Functional Linguistics: Halliday

language as a social semiotic
language as a stratified system

- A significant aspect of the meaning found in LLM-generated sequences appears to be due to generalisations reflecting such 'higher-order' patterns.
- It is not only that the sequences generated appear grammatical or formally correct, **many higher-order patterns** are unfolding within the generated responses as well.
- No arbitrary 'cut-off point' where we talk of 'form' on the one hand and 'meaning' on the other...



all strata are formal; all strata are functional

The semiotic nature of computer interfaces

“human-computer interaction is constituted to two separate processes: a sign process and a signal process. The user is heavily involved in the sign process. The computer maintains the signal process. Both run in parallel and **must be coupled**, more or less tightly, for interaction to take place.”

Frieder Nake 1994: 193 (emphasis added)

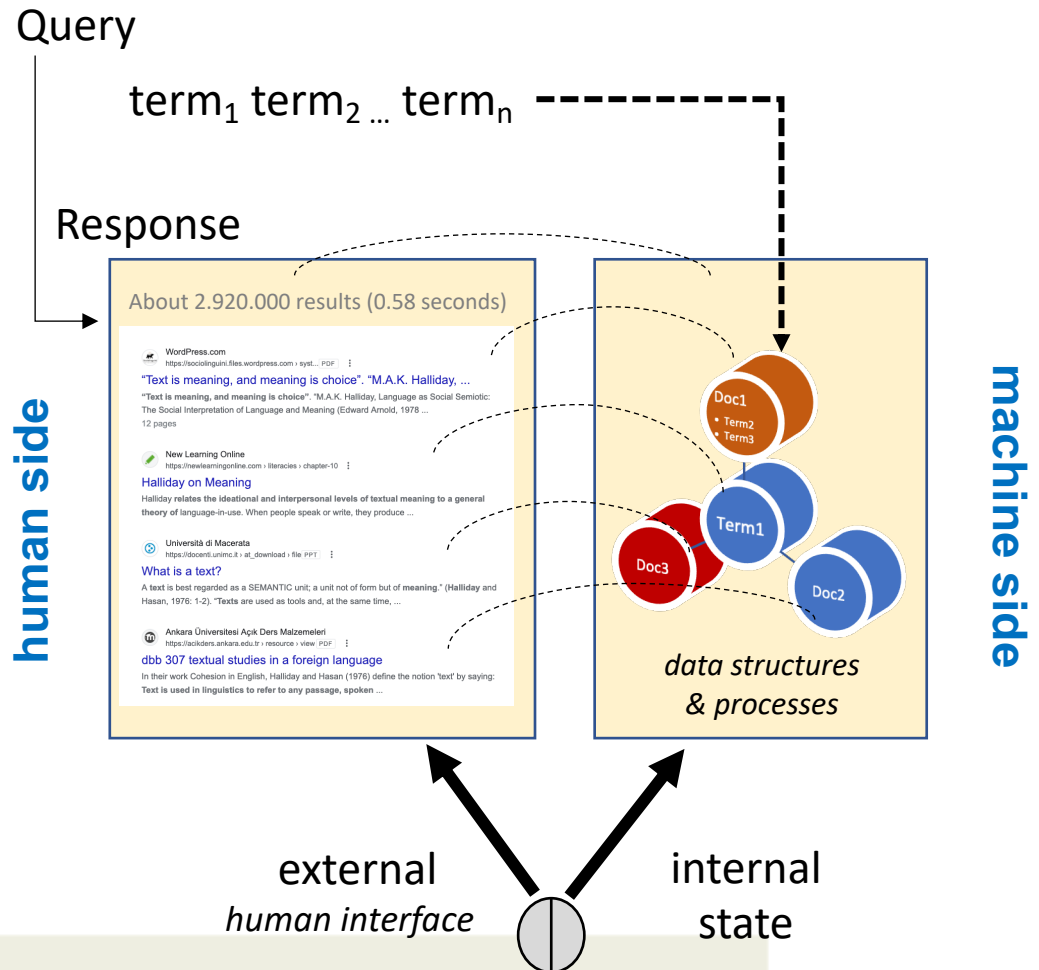


© Matej Meza / Universität Bremen

Coupling across sign and signal

HCI

designed congruence
between internal state and
external presentation

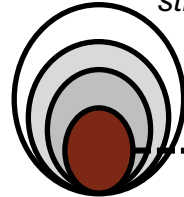




Non-coupling across sign and signal

Question

stratified semiotic

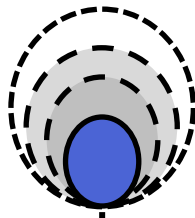


question formulated as text

Response

... In terms of the definition of text, Martin generally aligns with Halliday's perspective, emphasizing the functional and social aspects of language. However, Martin has delved into detailed analyses of different types of texts, exploring how language is used to construct meaning in specific contexts. ...

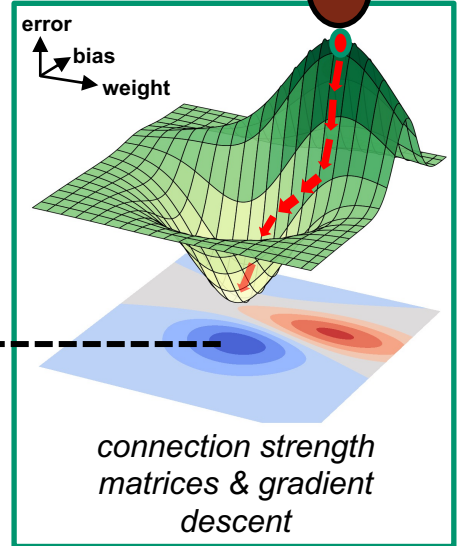
stratified semiotic



reconstruction

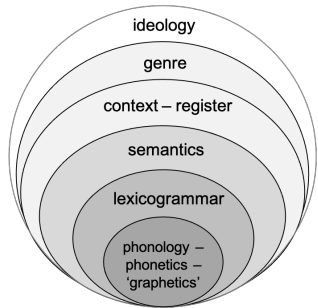
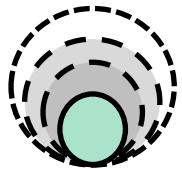
external human interface

human side



machine side

internal state



LLM-based Chat

designed **incongruence** between internal state and external presentation



The collapse of indexicality (again)

- Throughout language's existence: To the extent that certain chains of cohesive reference, certain style choices, certain generic structures were selected by the performance of the language model, those forms **were reliable indexes** for those choices, i.e., those choices were 'meant' by the system.
- What is more, there is often, from the output, **no way of telling** that particular instantiations are missing. This is effectively masked both by the training process strategising plausibility and again by computer interfaces that present results as if unproblematically complete from the *human* perspective.
- This is, therefore, by no means a question of LLM systems having 'no meaning', more it is a question of **which** meanings

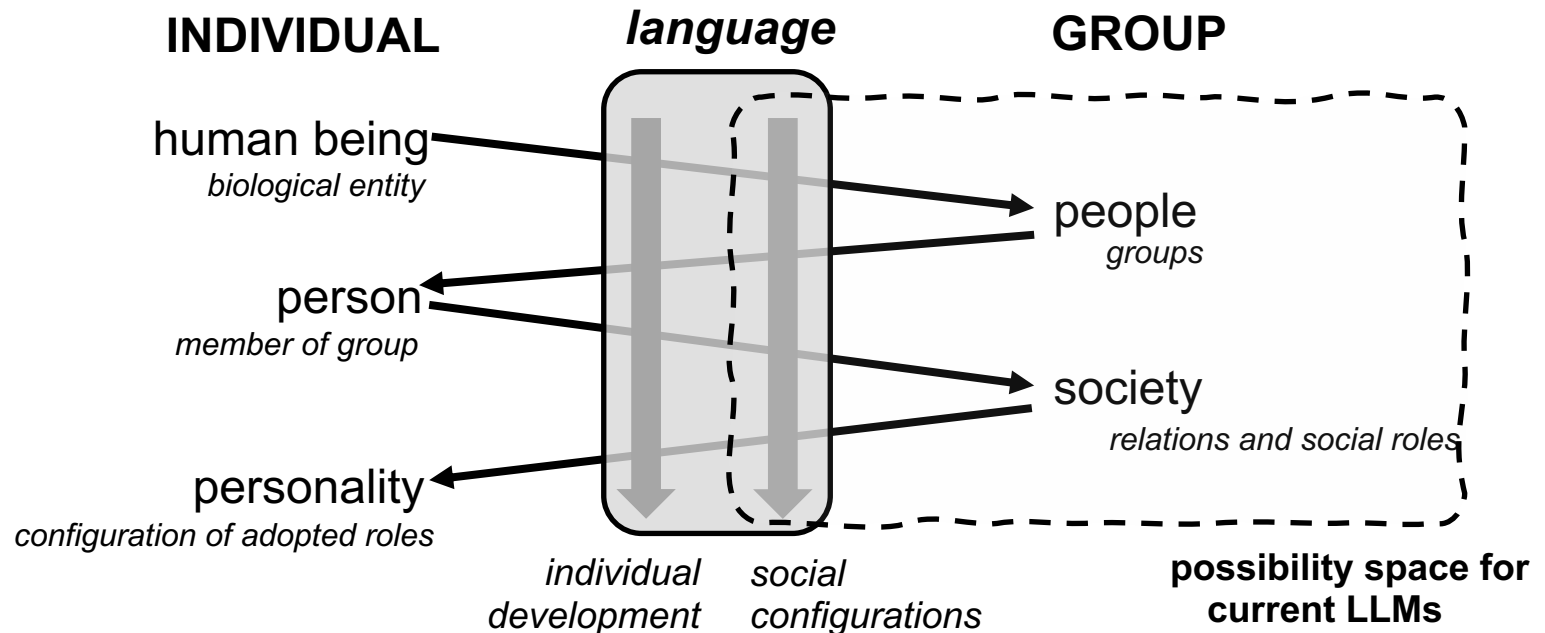
... and some serious consequences

- In fact, these produced sequences are more than just 'plausible'; they are **so** plausible (and plausible in particular ways) that an LLM **"can hijack our intuitive capacity to evaluate the accuracy of its outputs"** (Sobieszek & Price 2022: 341)
 - In other words: such outputs are not problematic because they have no meaning, **but because they have too much.**
- it is then not only that such output may then slip by our defence systems on plausibility and good judgement, their design is such that we will scarcely **have any choice in this.**
- ... by virtue of co-instantiation and indexicality

Semiotic Zombies?

Philosophical zombies (Kirk 1974)

the development of 'social beings' according to Halliday and SFL



Conclusion and Outlook

- theoretical message:
beware of semiotic zombies!
(but how??)



- a new direction for some practical/ongoing work:
probing for the kinds of information that have been learned
organised by strata

