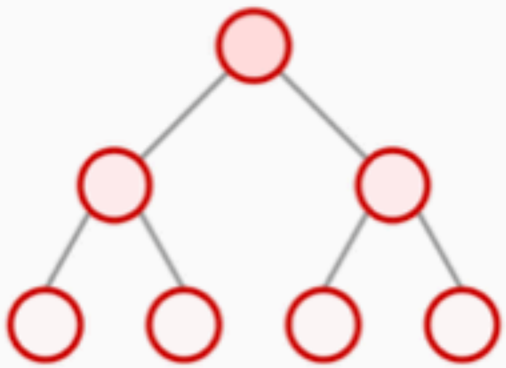
The Challenge of Compositionality for AI

June 29-30, 2022



A two-day online workshop on compositionality and artificial intelligence organized by Gary Marcus and Raphaël Millière.

What is compositionality?

A standard, theory-neutral way to state the principle of compositionality is as follows:

(C_0) The meaning of a complex expression is a function of the meanings of its constituents and the way they are combined.

Zoltán Gendler Szabó/ The case for compositionality

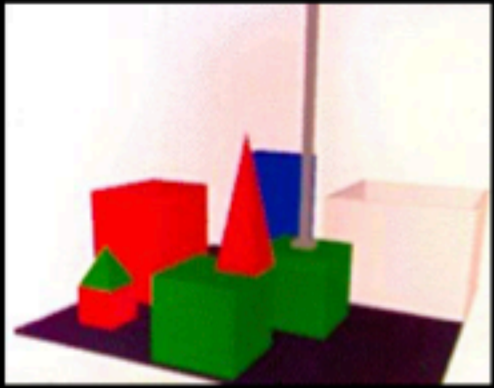
- *The cat is on the mat* specifies a meaning in terms of a relationship (ON) between some entities (THE CAT, THE MAT)
- *the mat is on the cat* specifies a *different* meaning; entities are same but relationship between entities is reversed
- Syntax *guides* the semantics
- Part of the **goal** of language comprehension is to recover those relationships
- Part of the goal of language production is to take an intent (specified in terms of relations between entities) in order to produce a structured string that represents the intent
- [also critical in vision, music, math, etc!]

Is compositionality optional?

- Large Language Models don't directly implement compositionality — at their peril
  - Whereas "semantic parsers" map sentences to meanings/intents, LLMs typically simply predict next words.
  - Those predictions are *correlated* with (traditional) meanings, but they aren't meanings.
    - there is no decomposition of a sentence into eg entities and relationships between those entities
    - and no accessible database that is (directly) updated
- This comes at a cost

What's at stake (1): Dynamically-updated world models

- In a classical framework (eg SHRLDU), you can relate a compositionally-composed utterance to a dynamically-updated database



Person: DOES THE SHORTEST THING THE TALLEST PYRAMID'S SUPPORT SUPPORTS SUPPORT ANYTHING GREEN?
Computer: YES, THE GREEN PYRAMID.
Person: WHAT COLOR IS IT?
Computer: BY "IT", I ASSUME YOU MEAN THE SHORTEST THING THE TALLEST PYRAMID'S SUPPORT SUPPORTS.
Computer: RED

- Harder to do that (maybe impossible?) with unanalyzed points in vector space

Thread

Gary Marcus @GaryMarcus · May 13

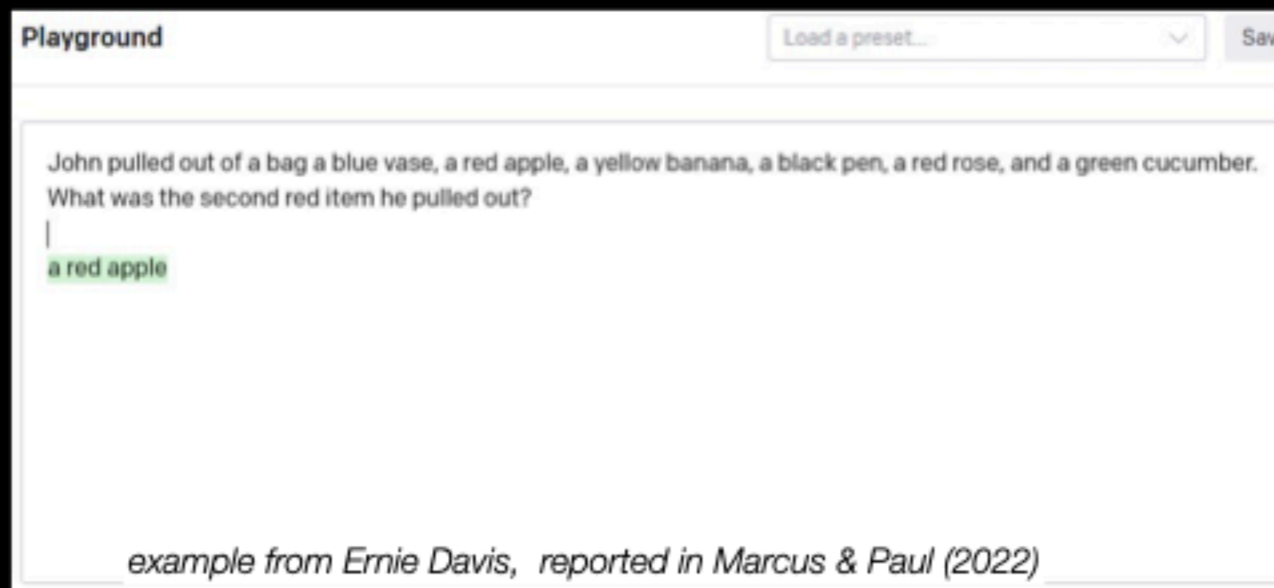
Folks @deepmind, have you tried to test #Gato on the complete set of materials in Winograd's classic 1968 SHRLDU Blocks World, in a robot arm plus image captioner + chat?

A full report on that would be fascinating!

@scott_e_reed @NandoDF @RaiaHadsell @OriolVinyalsML

Blocks World would likely pose challenges for Gato

- GPT in general loses context over time
- Indirect tests* indicate struggles with compositional interpretation
    - *Prediction per se is not directly interpretable, so proxies for meaning can only be assessed indirectly eg via word problems



example from Ernie Davis, reported in Marcus & Paul (2022)

What's at stake (2): Controllability

- pure prediction is hard to control; with holistic prediction, without interpretable meanings and database updates, you get terrific, broad linguistic coverage, BUT…
  - it's hard to ground LLMs ethically
    - tons of problems with bias and stereotyping, counseling harm etc
  - it's hard to ground them in terms of truth;
    - fabrication is frequent
  - it's hard to maintain coherence over the long term
- In systems like GPT-3, you can wind up a toxic spew of harmful advice and misinformation

DeepMind

Ethical and social risks of harm from Language Models

Abstract

Compositionality is not mysterious

- Programming languages assume it, for example. So does math.
    - The semantics of (eg) a Python program are determined by the parts and the ways in which those parts are put together
    - Programs are represented by syntactic trees that have semantics that can directly be inferred from those trees.
- What *is* mysterious:
    - The precise nature of compositionality in human language
    - The proper role and implementation of compositionality in AI

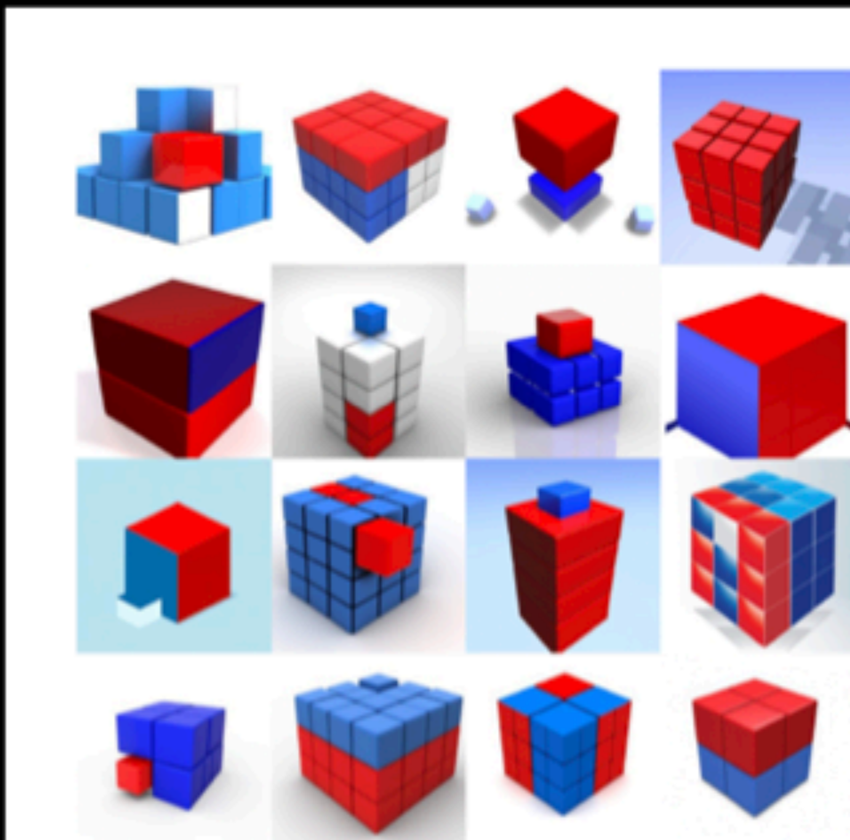
Three options going forward

- We could build all of our AI on a symbol-manipulation framework, in which compositionality is explicit, and well-understood.
  - Lots of potential virtues in terms of verifiability and intepretability
  - Symbol systems are typically largely hand-wired, often brittle, not entirely satisfactory
- We could ignore the issue of compositionality, and hope that with enough data, things will sort themselves out.
  - LLM's produce strings that reflect the grammar of human language
  - But lack stability and grounding and do not produce interpretable meanings of input language
- We could try to find ways of incorporating compositionality into neural networks
  - Smolensky (1988, 2022); Marcus (2001)

Banking on scaling alone might not be the best strategy

A look at compositionality in DALL-E
— with examples of what you might hope for, and how it fails

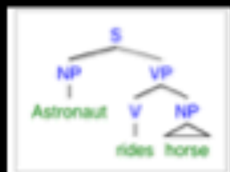
Dall-E 2 has lots of data, and lots of problems w compositionality

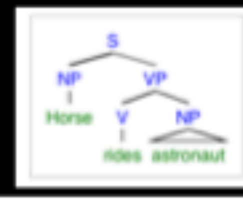


"a red cube on top of a blue cube".

Marcus, Davis, Aaronson (2022, arxiv)

Horse Rides Astronaut

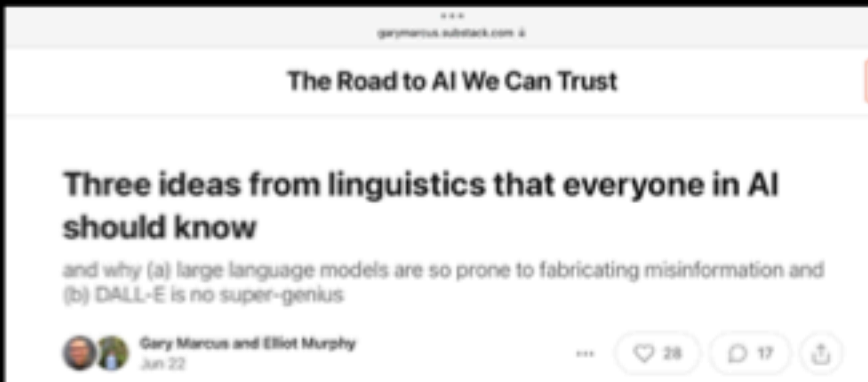




A horse riding an astronaut.

https://garymarcus.substack.com/p/horse-rides-astronaut

comparatives



garymarcus.substack.com

The Road to AI We Can Trust

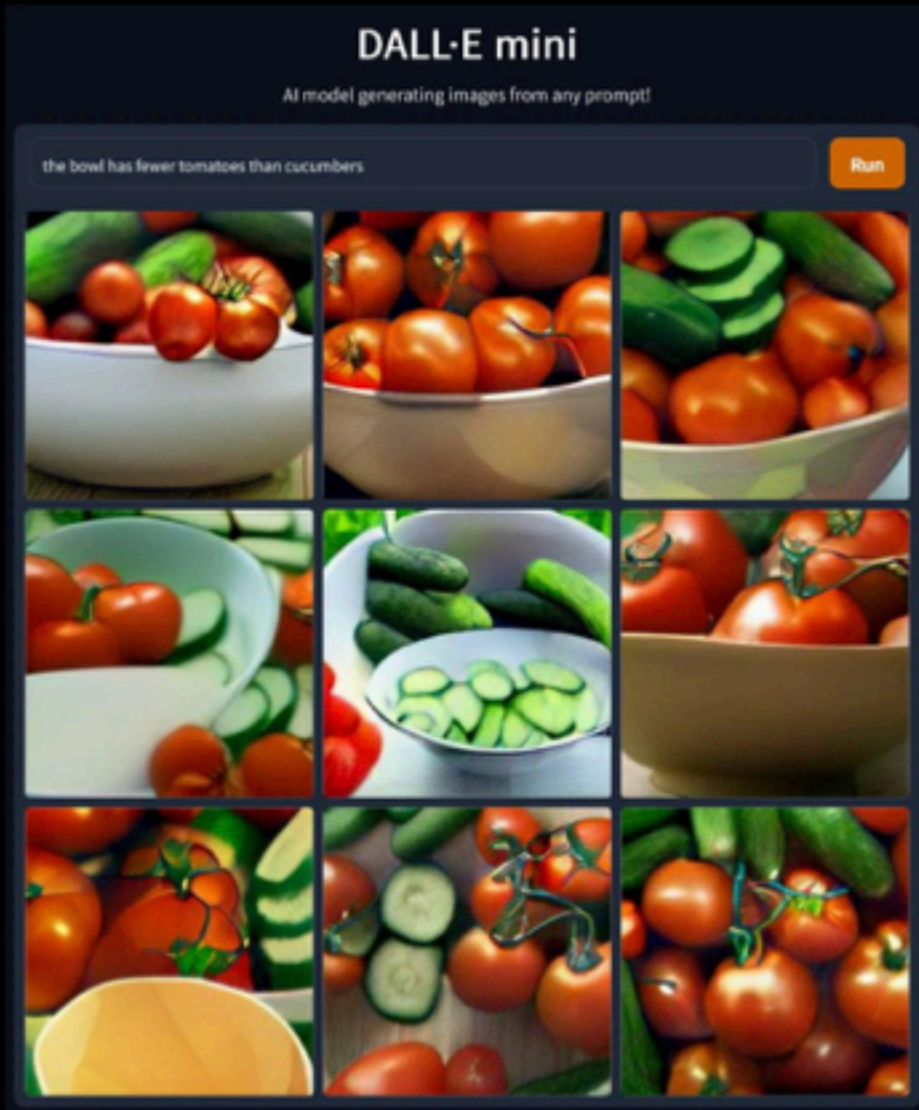
Three ideas from linguistics that everyone in AI should know

and why (a) large language models are so prone to fabricating misinformation and (b) DALL-E is no super-genius

Gary Marcus and Elliot Murphy
Jun 22

28 17

and see forthcoming manuscript, with Evelina Leivada



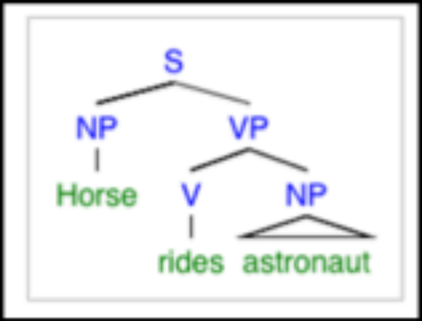
Does this mean neural networks are *incompatible* with compositionality?

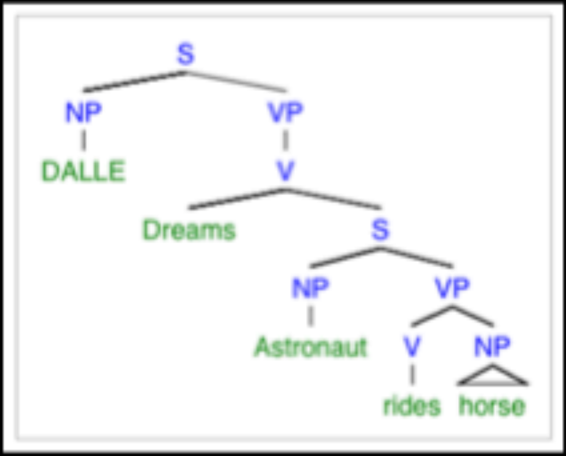
- **No!**
- It means that *some* neural networks (e.g. DALL-E 2) are incompatible.
- Any symbol-manipulating system can be implemented (realized) in many different ways, including in a neural network (McCulloch & Pitts; Siegalmann & Sontag).
- The real question is, if we are to build AI out of neural networks, must we build a neural network that *implements* compositionality, in a way that **maps** 1:1 onto a classic symbolic system, or might a successful neural network offer some kind of alternative? (What kind of alternative?)
- Subquestion: might compositionality be something that is *learned*, rather than being something inherent (a la in the design of LISP)?

Minimal Requirements for Compositionality

as developed e.g. in Marcus (2001)

- Stable encodings of individual elements
- An operation that concatenates pieces of trees together
  - or disassembles wholes into parts
- Iterative processes for (de)constructing larger structures
- Representational formats for trees (or something very similar)





- Plus: Linking mechanisms that derive semantics relative to syntactic representation (eg Pinker 1984, 1989)

the thing that Hinton is trying to do is very relevant

- GLOM is really an effort at building stable encodings that could be used in representations of complex wholes, very much like slide 1 on previous slide

arXiv > cs > arXiv:2102.12627

Search...

Help | Adva

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 25 Feb 2021]

How to represent part-whole hierarchies in a neural network

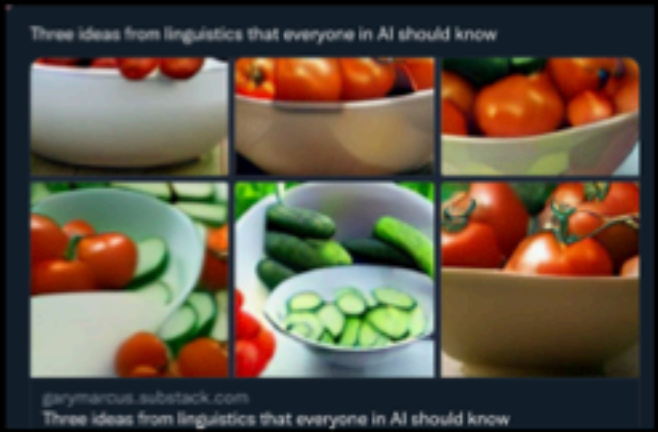
Geoffrey Hinton

This paper does not describe a working system. Instead, it presents a single idea about representation which allows advances made by several different groups to be combined into an imaginary system called GLOM. The advances include transformers, neural fields, contrastive representation learning, distillation and capsules. GLOM answers the question: How can a neural network with a fixed architecture parse an image into a part-whole hierarchy which has a different structure for each image? The idea is simply to use islands of identical vectors to represent the nodes in the parse tree. If GLOM can be made to work, it should significantly improve the interpretability of the representations produced by transformer-like systems when applied to vision or language

Comments: 43 pages, 5 figures

Compositionality is *not* sufficient; it is a part of a framework

- Syntax -> Semantics -> cognitive models [best guess at external world, fictional worlds]
- we use language to *accumulate* knowledge ("is junk food more or less expensive than regular food?", "do people make junk food? do they grow it?")
- The real challenge is to build language understanding systems that can update their understanding of the world by decomposing meanings in terms of their parts, taken in context of speaker intent.



A few words about humans

- Humans are interesting; we clearly understand wholes in terms of their parts, but there are also some deviations from ideal.
  - Machines allow arbitrary embedding; humans have trouble with center embedding (*A man that a woman that a child that a bird that I heard saw knows loves*)
    - My view: variable binding is expensive in humans, and we use a cue-dependent substitute that is vulnerable to inference (Marcus, 2008).
  - Humans allow an immense number of “frozen forms” and idioms that are not internally compositional (*kick the bucket, dead end*, etc).

Idioms are part of what makes NLU hard

- You don’t understand *kick the bucket* by forming a representation of someone sending a projective force towards a pail.
- A good NLU system must blend (at least) two pathways:
  - pure semantics from syntax (which works for *tipped over the pail*)
  - with idiomatic retrieval (*kick the bucket = died)*
- A single sentence can combine both:
  - The person who tipped over the pail on Tuesday suddenly and unexpectedly kicked the bucket on Wednesday)
  - Getting all this right cries out for ML and classical NLU to work together

Conclusions

- Compositionality in language is about systematically inferring (or generating) a meaning from parts, in a structure-dependent way
  - Flows naturally in symbolic paradigms (e.g. Python has a clear, structure-dependent semantics)
  - It doesn't automatically emerge from very large data (DALL-E)
  - You need *some* kind of innate architectural underpinning.
- No fully adequate solution exists
  - Hand-writing all rules of a language is difficult
  - There is a large idiomatic periphery that ML ought to be able to help with
  - Current ML approaches tend to focus on feature-wise correlations; we need robust ML that works at scale over higher level abstractions
  - Hence lots of reasons for a r'approchment between symbolic and statistical approaches
- Compositionality ultimately is in service of something larger: dynamically updated cognitive models of the world. Capturing that workflow is vital if we are to build systems we can trust.