Emergent Properties

20 Oct 2023
FSU SC-ML
Tom Juzek

Emergent Properties

20 Oct 2023
FSU SC-ML
Tom Juzek
+Nathan

Resources General

Wei et al. 2022: https://arxiv.org/pdf/2206.07682.pdf

Steinhardt 2022: https://bounded-regret.ghost.io/future-ml-systems-will-be-qualitatively-different/

Anderson 1972: https://scholar.google.com/scholar?cluster=3979541075862383168&hl=nl&as_sdt=0,10

Power et al. 2021: https://mathai-iclr.github.io/papers/papers/MATHAI_29_paper.pdf?ref=bounded-regret.ghost.io

Schaeffer et al. 2023: https://arxiv.org/pdf/2304.15004.pdf

Resources Linguistics

Warstadt and Bowman 2022:
https://arxiv.org/pdf/2208.07998.pdf

Warstadt et al. 2019:
https://arxiv.org/pdf/1805.12471.pdf

Anon SAD 2023 (paper+corpus*+code*):
https://github.com/arizus/sad

Overview

- Intro, definition, background
- deeper dive, examples, challenges
- linguistic properties, current research line

→ usual format: Qs anytime

A very general introduction

Data - ODEP – ANN – Transformers – LLMs

A very general introduction

“AI/LLM revolution”, “structure of learning”

A very general introduction

→ Emergence

Definition

“Emergence is when quantitative changes in a system result in qualitative changes in behavior.”

Steinhardt 2022, “rooted in” (as per Wei et al. 2022):

Definition

“... we have begun to formulate a general theory of just how this shift from quantitative to qualitative differentiation takes place. This formulation, called the theory of "broken symmetry," may be of help in making more generally clear the breakdown of the constructionist converse of reductionism”

Anderson 1972

Definition

“Emergence is when quantitative changes in a system result in qualitative changes in behavior.”

https://bounded-regret.ghost.io/future-ml-systems-will-be-qualitatively-different/

water?

Definition

“Emergence is when quantitative changes in a system result in qualitative changes in behavior.”

sharp & unpredictable

Definition

ChatGPT is not entirely happy with this (“generally good but not quite right”).

“Emergence in machine learning refers to the phenomenon where a model develops complex behaviors or attributes not explicitly engineered into its components, arising from intricate interactions between its learned parameters during training.”

Definition

ChatGPT’s definition of behaviour (which is pretty much wiki…):

“Behavior refers to the range of actions and mannerisms exhibited by entities—organisms, systems, or artificial agents—in response to their environment or stimuli. This includes the physical movements, responses, and thought processes that are provoked by, or interact with, external or internal cues.”

Definition

ChatGPT’s formal definition:

…

Definition

1. **System Description:**

   Let's consider a system S consisting of n components, where each component i can be described by a variable or set of variables x_i. The state of the system at time t can be represented as a vector $X(t) = [x_1(t), x_2(t), ..., x_n(t)]$.

2. **Interaction Among Components:**

   The interactions among the components can be represented by a function F, which maps the current state of the system to a new state. In a dynamic system, this could be represented as:

$$\frac{dX(t)}{dt} = F(X(t))$$

   In the context of deep learning, F could represent the architecture of the neural network (including the activation functions, structure, etc.), and $X(t)$ could represent the states of the neurons.

Definition

3. **Quantitative to Qualitative Change (Emergence):**

   Emergence happens when a quantitative change leads to a qualitative change. Mathematically, we are looking at a transformation T that would represent this non-linear shift:

$$Y(t) = T(X(t))$$

   where $Y(t)$ represents the new qualitative properties of the system. The transformation T is not merely a linear transformation but a complex, often non-deterministic function that maps the quantitative changes in X to qualitative changes in Y.

   In deep learning, this could represent the point at which simple adjustments in weights during training (quantitative) suddenly result in a neural network gaining a markedly improved capability, like image recognition or language processing (qualitative).

Definition

4. **Measurement of Emergence:**

   Defining a measure E for emergence is highly abstract and depends on the system's context. However, it could be considered as a function of the difference between the behavior of the whole system versus the sum of behaviors of its parts.

$$E = G(Y(t)) - \sum_{i=1}^{n} g(x_i(t))$$

   Here, G is a function measuring the total system behavior, and g measures individual component behaviors. E is then a measure of how the collective behavior is not just the sum of parts, representing the novel properties or behaviors emerging from the system.

General Relevance

Emergence of *something* before our eyes:

(GPT-1) → GPT-2 → HF “GPT 2.5” → GPT-3 → …

(All transformers)

Architectural changes

‘100 tricks stacking up’

Model size ~ Training data size

General Relevance

Emergence of *something* before our eyes:

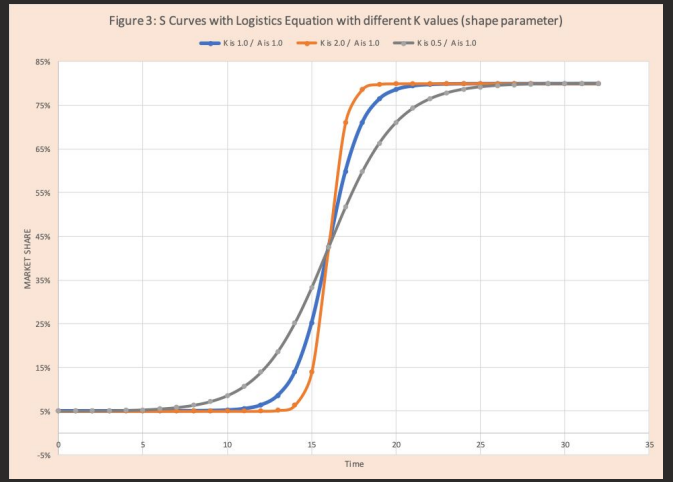
(GPT-1) → GPT-2 → HF “GPT 2.5” → GPT-3 → …

(All transformers)

Architectural changes

‘100 tricks stacking up’

Model size ~ Training data size



Product adoption, K Fordyce

General Relevance

Emergence of *something* before our eyes:

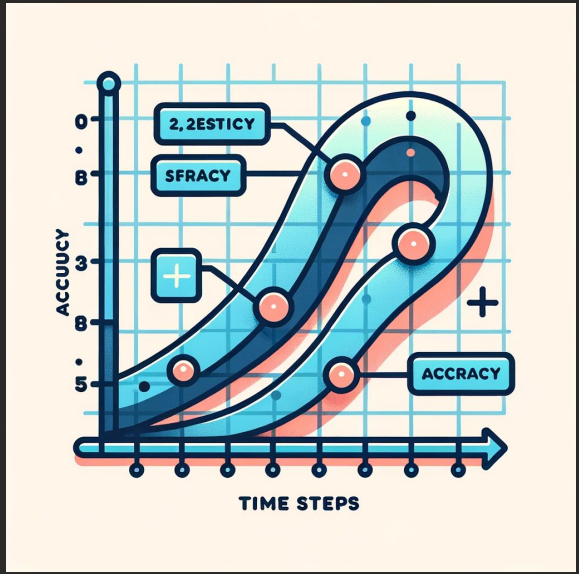
(GPT-1) → GPT-2 → HF “GPT 2.5” → GPT-3 → …

(All transformers)

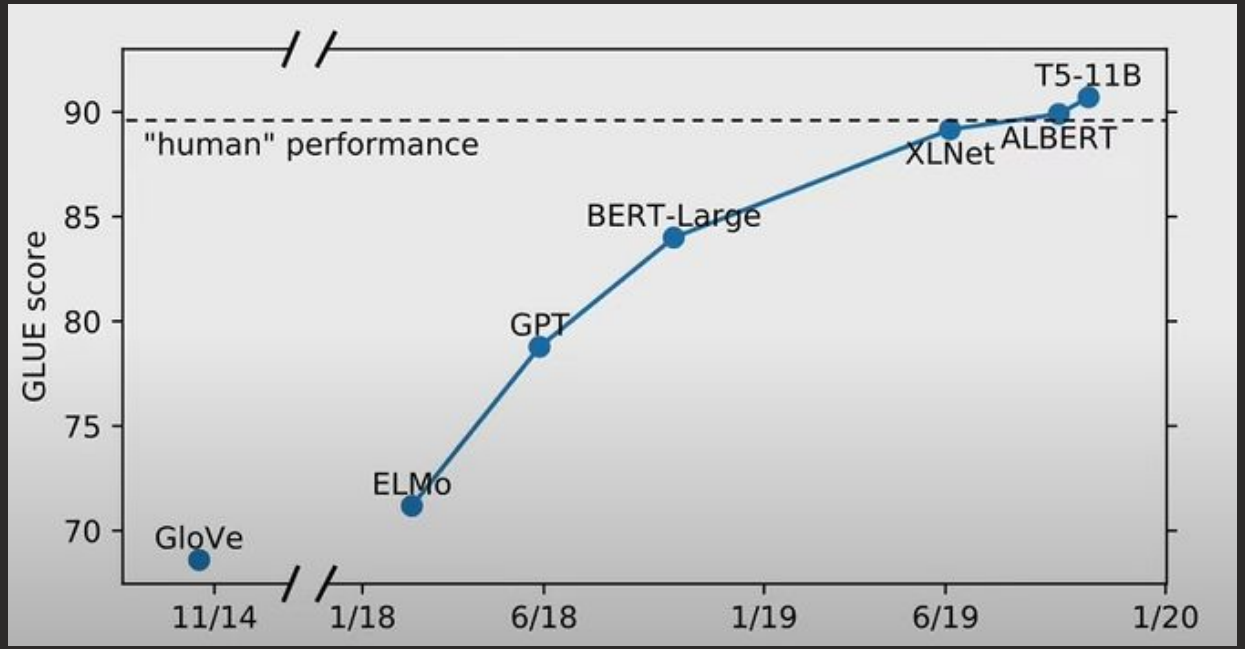
Architectural changes

‘100 tricks stacking up’

Model size ~ Training data size



Examples: GLUE



https://machinehack.com/story/large-language-models-llm-and-its-future-of-ai

General Relevance

AI emergence

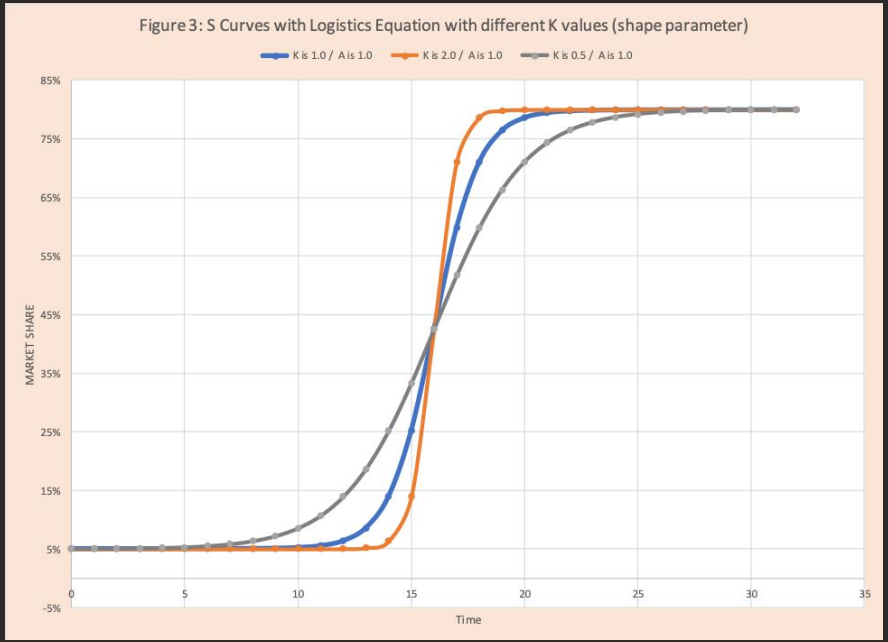
vs

Emergence in specific tasks

→ sharpness & change in quality

Surface modelling: Sharpness

We have touched on the S-curve



Surface modelling

```
x (float): The input value for which the function is calculated.

L (float): The maximum value of the curve.

k (float): The steepness of the curve.

x0 (float): The x-value of the sigmoid's midpoint.
```

```
return L / (1 + math.exp(-k * (x - x0)))
```

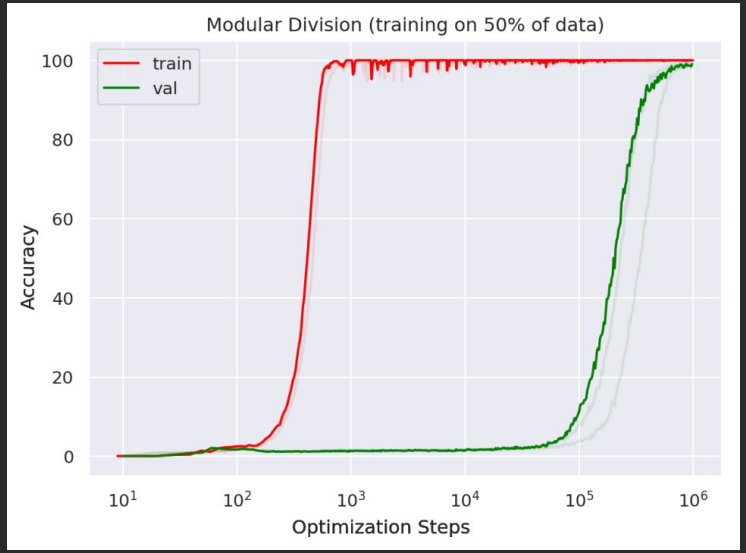
→ surface modelling of (most of) what we are seeing

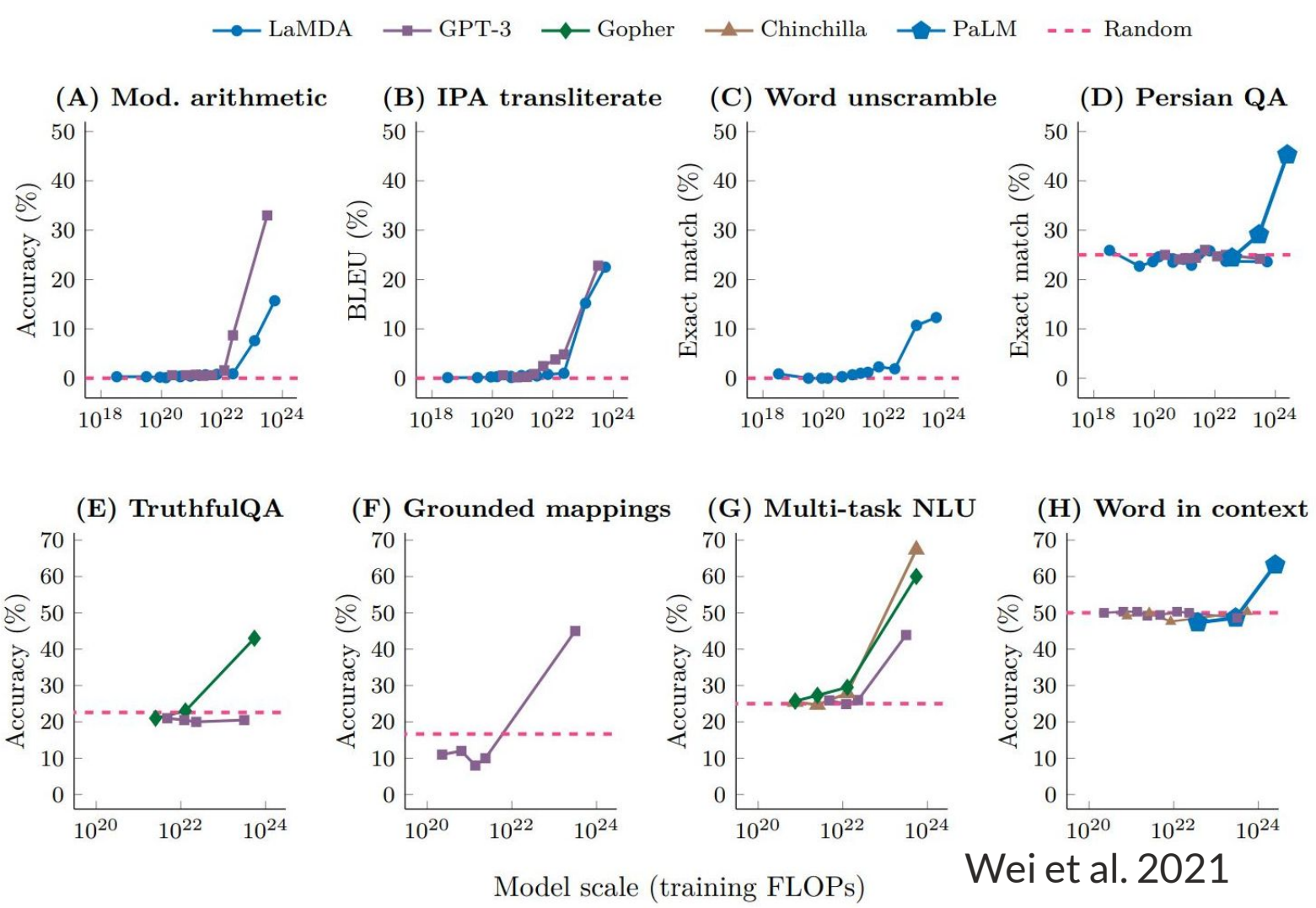
Examples

Similar developments is what researchers have been observing for various TASKS

Examples

Power et al. 2021: Puzzle solving, “grokking” (delayed val onset),





Wei et al. 2021

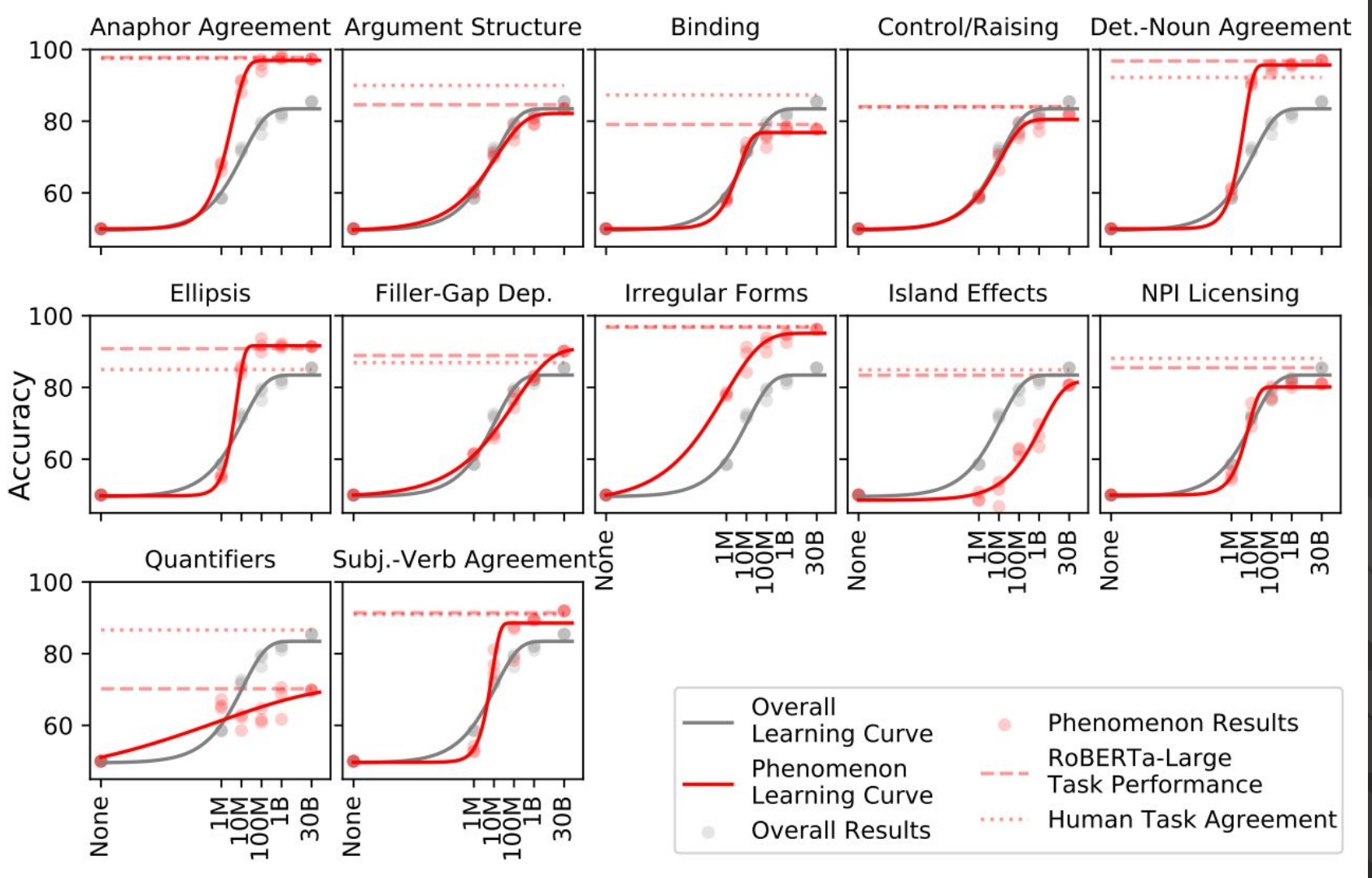
Syntactic tasks

Minimal pairs of various syntactic constructions.

1) `Mary asks John to come home.`
2) `Mary tries John to come home.`

Choose the correct one!

(Warstadt & Bowman 2022)


Anaphor Agreement
Argument Structure
Binding
Control/Raising
Det.-Noun Agreement
Ellipsis
Filler-Gap Dep.
Irregular Forms
Island Effects
NPI Licensing
Quantifiers
Subj.-Verb Agreement
Accuracy
100
80
60
None
1M
10M
100M
1B
30B
Overall Learning Curve
Phenomenon Learning Curve
Overall Results
Phenomenon Results
RoBERTa-Large Task Performance
Human Task Agreement

Bridge

“What makes emergent abilities intriguing is two-fold: their sharpness, transitioning seemingly instantaneously from not present to present, and their unpredictability, appearing at seemingly unforeseeable model scales.”

Schaeffer et al. 2023

Definition

“Emergence is when quantitative changes in a system result in qualitative changes in behavior.”

cars to traffic

Big Qs

“What controls which abilities will emerge? What controls when abilities will emerge? How can we make desirable abilities emerge faster, and ensure undesirable abilities never emerge?”

Schaeffer et al. 2023

Critical voices

There is no qualitative change

It's always cars

→ General AI context: earlier hint: smaller models with high quality data doing as well as big models (*textbooks are all you need*)

→ For the tasks: …

Schaeffer et al. 2023

“ [W]e present an alternative explanation for emergent abilities: that for a particular task and model family, when analyzing fixed model outputs, emergent abilities appear due the researcher’s choice of metric rather than due to fundamental changes in model behavior with scale.”

to be continued
27 Oct 2023

Thank you