

CNRS French–Danish Workshop
Reasoning in the Embedding Space
Copenhagen, Denmark

Empiricism vs. Formalism

In the Study of Embeddings

Juan Luis Gastaldi

`www.giannigastaldi.com`

ETH zürich

January 20, 2026

Empiricist Turn in Computer Science

Chomsky's Trap

Language Models as Formal Objects

Philosophical Consequences

Takeaways

Empiricist Turn in Computer Science

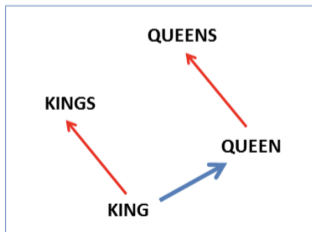
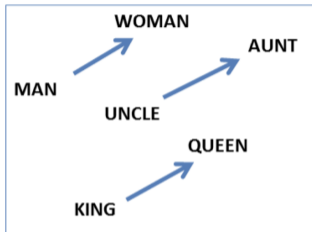
Chomsky's Trap

Language Models as Formal Objects

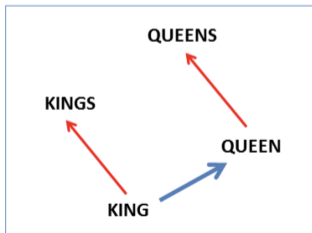
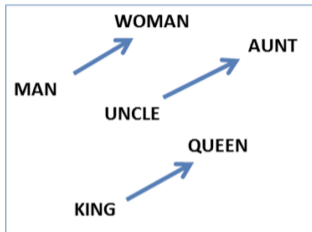
Philosophical Consequences

Takeaways

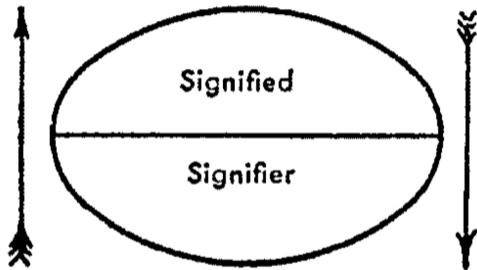
Anna's



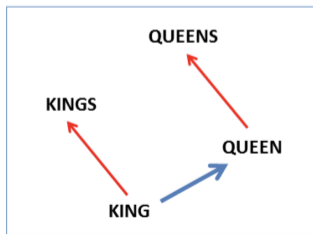
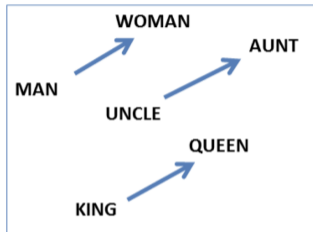
Anna's



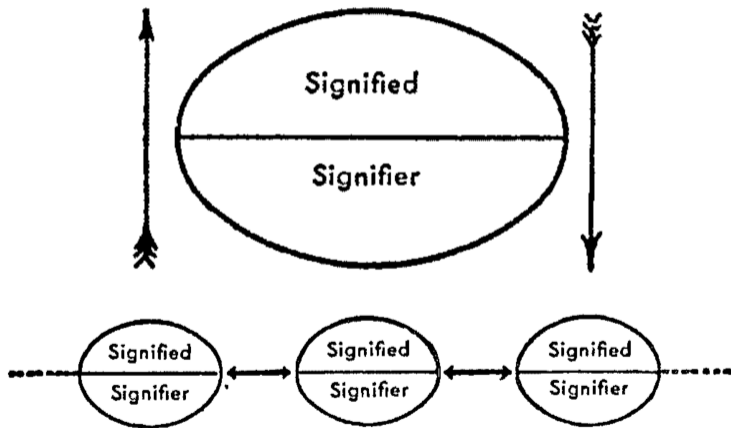
Gianni's



Anna's

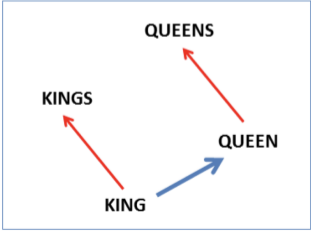
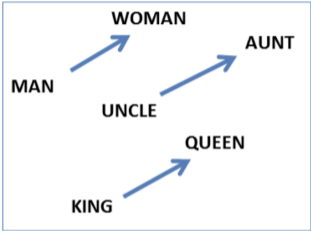


Gianni's



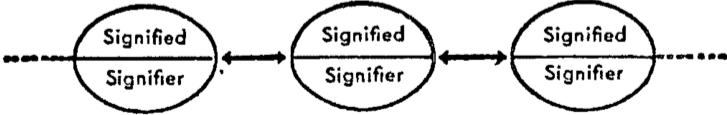
Teen Room Posters

Anna's

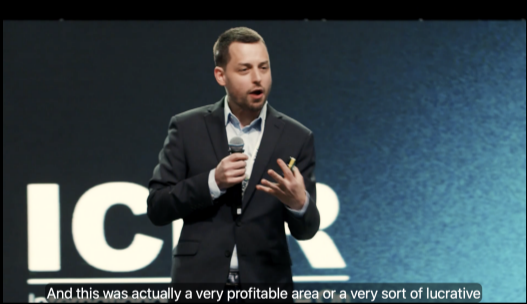


Gianni's

We then have a function between a as one terminal and the functional field *sovereign* φ *man* as the other, and likewise a function between b as one terminal and the functional field *sovereign* φ *woman* as the other. Diagrammatically:



Empirical Saturnalia



And this was actually a very profitable area or a very sort of lucrative

The empirics of deep learning

(Circa 2020) the scaling era is here; deep networks are now just emergent things we have created, that have to be studied scientifically like any other physical phenomenon

It seemed like **the best way for academic research to influence the field** is to develop the biology/physics (and let's be honest, more often pop psychology) of existing large models

24

Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

Can Large Language Models Be an Alternative to Human Evaluation?

Cheng-Han Chiang
National Taiwan University,
Taiwan
dcm10714@gmail.com

Hung-yi Lee
National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

And this was actually a very profitable area or a very sort of lucrative

The empirics of deep learning

(Circa 2020) the scaling era is here; deep networks are now just emergent things we have created, that have to be studied scientifically like any other physical phenomenon

It seemed like **the best way for academic research to influence the field** is to develop the biology/physics (and let's be honest, more often pop psychology) of existing large models

24

Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

DO LLMs HAVE CONSISTENT VALUES?

Naama Rozen

Tel-Aviv University
naamarozen240@gmail.com

Liat Bezalel

Tel-Aviv University
liatbezalel@mail.tau.ac.il

Gal Elidan

Google Research
Hebrew University
elidan@google.com

Amir Globerson

Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel

Tel-Aviv University
della@tauex.tau.ac.il

Cheng-Han Chiang

National Taiwan University,
Taiwan
dcm10714@gmail.com

Hung-yi Lee

National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

The empirics of deep learning

ca 2020) the scaling era is here; deep networks are now just emergent things we have created, that have to be studied scientifically like any other physical phenomenon

seemed like **the best way for academic research to influence the field** is to develop the biology/physics (and let's be honest, more often pop psychology) of existing large models

24

And this was actually a very profitable area or a very sort of lucrative

Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

DO LLMs HAVE CONSCIOUSNESS?

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Cheng-Han Chiang
National Taiwan University,
Taiwan
dcm10714@gmail.com

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Recursive Self-Improvement through Reinforcement Learning

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

Taro Yano
NEC
taro_yano@nec.com

Masafumi Oyamada
NEC
oyamada@nec.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau.ac.il

Hung-yi Lee
National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

deep learning

(ca 2020) the scaling era is here, deep networks are now just emergent things
we have created, that have to be studied scientifically like any other physical
phenomenon

emerged like **the best way for academic research to influence the field** is to
develop the biology/physics (and let's be honest, more often pop psychology) of
existing large models

And this was actually a very profitable area or a very sort of lucrative

24

Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

DO LLMs HAVE CONSCIOUSNESS?

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Reinforcing Self-Improvement through

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau.ac.il

DO LLMs “KNOW” INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Juyeon Heo^{1*} Christina Heinze-Deml² Oussama Elachqar² Kwan Ho Ryan Chan^{3*} Shirley Ren² Udhay Nallasamy² Andy Miller² Jaya Narain²
¹University of Cambridge ²Apple ³University of Pennsylvania
jh2324@cam.ac.uk jnarain@apple.com

Cheng-Han Chiang
National Taiwan University,
Taiwan
dcm10714@gmail.com

Hung-yi Lee
National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

...emerged like **the best way for academic research to influence the field** is to develop the biology/physics (and let's be honest, more often pop psychology) of existing large models

Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

Empirical Saturnalia

DO LLMs HAVE CONSCIOUSNESS?

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Cheng-Han Chiang
National Taiwan University,
Taiwan
dcm10714@gmail.com

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Domain-Specific Self-Improvement through Reinforcement Learning

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau.ac.il

Hung-yi Lee
National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

DO LLMs “KNOW” INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Juyeon Han¹, Gyeongmin Hong², Dongmin Lee², Gwanjoon Eom², Kwanghyun Park², Seungyeon Park², Udhay Narayanan¹
¹University of Michigan, ²Google Research
jhan2324@umich.edu

DO LLMs RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMs

Siyan Zhao^{2*}, Mingyi Hong^{1,3}, Yang Liu¹, Devamanyu Hazarika¹, Kaixiang Lin¹ †
¹Amazon AGI, ²UCLA, ³University of Minnesota
siyanz@cs.ucla.edu, mhong@umn.edu, devamanyu@u.nus.edu
{yangliud, kaixianl}@amazon.com

24

And this was actually a very profitable area or a very sort of lucrative

Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

DO LLMs HAVE CONSCIOUSNESS?

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Cheng-Han Chiang
National Taiwan University,
Taiwan
dcml0714@gmail.com

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Domain-Specific Self-Improvement through Reinforcement Learning

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau.ac.il

Hung-yi Lee
National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

DO LLMs “KNOW” INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Juyeon Hwang*, Gihun Hwang*, Dongmin Gwon, Eunsol Kim, Hyeonjun Park, Seungjun Park, Seungyeon Park, Udhay Narayanan
¹University of Wisconsin-Madison
jh2324@wisc.edu

DO LLMs RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMs

Language Models are Few-Shot Learners

LLMs Are Not Intelligent Thinkers: Introducing Mathematical Topic Tree Benchmark for Comprehensive Evaluation of LLMs

Arash Gholami Davoodi¹, Seyed Pouyan Mousavi Davoudi, Pouya Pezeshkpour²
¹Carnegie Mellon University, ²Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Kyle Ryder* **Melanie Subbiah***
Pranav Shyam **Girish Sastry**
Gretchen Krueger **Tom Henighan**

And this was actually a very p...

DO LLMs HAVE CONSCIOUSNESS?

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Reinforcement Learning

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau.ac.il

DO LLMs “KNOW” INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Juyeon Hwang*, Gihun Hong*, Dongmin Gwon*, Eunsol Kim*, Hanmin Lee*, Seungjae Park*,
Udhay Narayanan*,
*University of Wisconsin-Madison
jh2324@wisc.edu

DO LLMs RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMs

Cheng-Han Chiang
National Taiwan University,
Taiwan

Hung-yi Lee
National Taiwan University,
Taiwan

When Can LLMs *Actually* Correct Their Own Mistakes? A Critical Survey of Self-Correction of LLMs

Ryo Kamoi¹ Yusen Zhang¹ Nan Zhang¹ Jiawei Han² Rui Zhang¹
¹Penn State University, USA ²University of Illinois Urbana-Champaign, USA
{ryokamoi, rmz5227}@psu.edu

Language Models are Few-Shot Learners

Scaling Mathematical Topic Tree Evaluation of LLMs

Mehdi Davoudi, Pouya Pezeshkpour²
Megagon Labs

mehdidavoudi@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Kyle Ryder* Melanie Subbiah*
Pranav Shyam Girish Sastry
Gretchen Krueger Tom Henighan

DO LLMs HAVE CONSCIOUSNESS?

Naama Rozen
Tel Aviv University

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Reinforcement Learning Self-Improvement through

Yoichi Ishibashi
NEC

DO LLMs “KNOW” INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Large Language Models are Zero-Shot Reasoners

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu
Google Research, Brain Team

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke Iwasawa
The University of Tokyo

Ryo Kamoi¹ Yusen Zhang¹ Nan Zhang¹ Jiawei Han² Rui Zhang¹
¹Penn State University, USA ²University of Illinois Urbana-Champaign, USA
{ryokamoi, rmz5227}@psu.edu

DO LLMs RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMs

Language Models are Few-Shot Learners

Scaling Mathematical Topic Tree Evaluation of LLMs

Ali Davoudi, Pouya Pezeshkpour²
Megagon Labs

alidavoudi@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Jack Ryder* Melanie Subbiah*

Pranav Shyam Girish Sastry

Gretchen Krueger Tom Henighan

Can Large Language Models Invent Algorithms to Improve Themselves?:
Algorithm Discovery for Reinforcement Learning through Self-Improvement through

DO LLMs HAVE CONSCIOUSNESS?

Sparks of Artificial General Intelligence:
Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehcke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

HOW DO LLMs "THINK" INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

DO LLMs RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING INSTRUCTIONS

Takeshi Kojima

The University of Tokyo

t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu

Google Research, Brain Team

Machel Reid

Google Research*

Yutaka Matsuo

The University of Tokyo

Yusuke Iwasawa

The University of Tokyo

Language Models are Few-Shot Learners

Scaling Mathematical Topic Tree
Evaluation of LLMs

Ryo Kamoi¹ Yusen Zhang¹ Nan Zhang¹ Jiawei Han² Rui Zhang¹

¹Penn State University, USA ²University of Illinois Urbana-Champaign, USA

{ryokamoi, rmz5227}@psu.edu

Abhinav Shrivastava, Pavan Kumar, Davoudi, Pouya Pezeshkpour²
Megagon Labs

abhinav@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Jack Ryder*

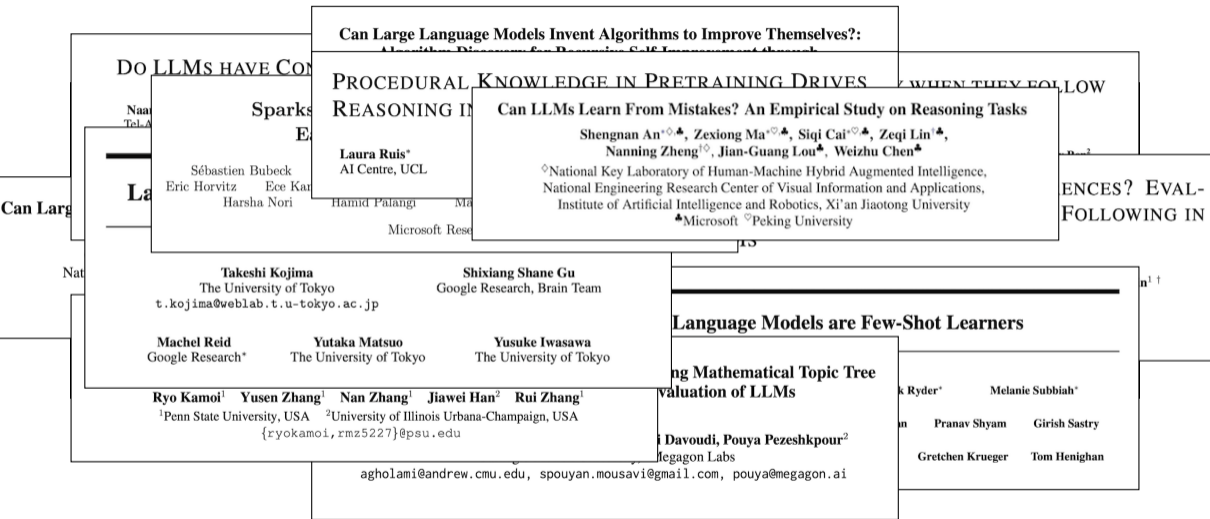
Melanie Subbiah*

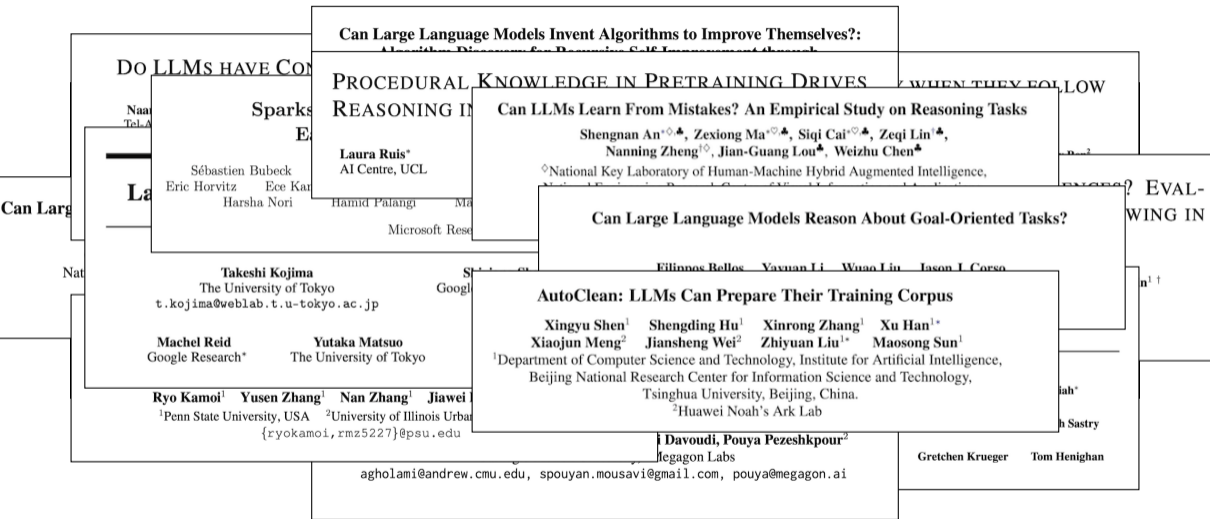
Pranav Shyam

Girish Sastry

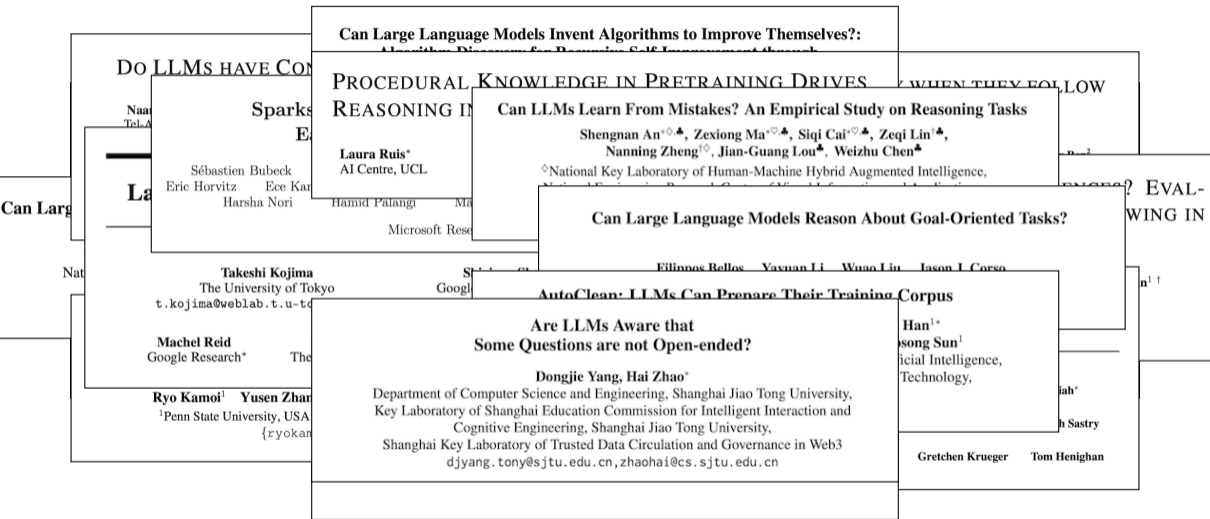
Gretchen Krueger

Tom Henighan





Empirical Saturnalia



Can Large Language Models Invent Algorithms to Improve Themselves?:
A Study in Recursive Self-Improvement (Abstract)

DO LLMs HAVE CONSCIOUSNESS?

PROCEDURAL KNOWLEDGE IN PRETRAINING DRIVES REASONING IN

WHEN THEY FOLLOW

Sparks

REASONING IN

Can LLMs Learn From Mistakes? An Empirical Study on Reasoning Tasks

Shengnan An^{*,†,‡}, Zexiong Ma^{*,†,‡}, Siqi Cai^{*,†,‡}, Zeqi Lin^{*,‡},
Nanning Zheng^{†,‡}, Jian-Guang Lou^{*,‡}, Weizhu Chen^{*,‡}

Laura Ruis^{*}
AI Centre, UCL

[†]National Key Laboratory of Human-Machine Hybrid Augmented Intelligence,

Naam
Tel: A

Sébastien Bubeck
Eric Horvitz Ece Kar

La

Self-Interpretability: LLMs Can Describe Complex Internal Processes that Drive Their Decisions, and Improve with Training

Dillon Plunkett
Northeastern University
d.plunkett@northeastern.edu

Adam Morris
Princeton University
thatadamorris@gmail.com

Keerthi Reddy
Independent Researcher

Jorge Morales
Northeastern University

Can Large Language Models Reason About Goal-Oriented Tasks?

Yayuan Li Wuzao Liu Jason I. Corso

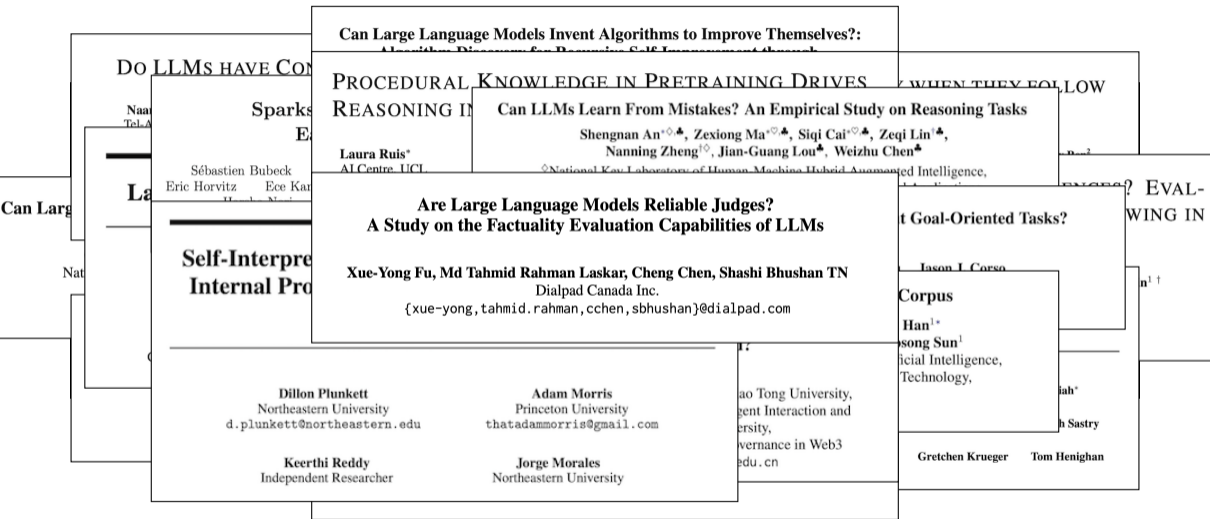
Can Large Language Models Reason About Goal-Oriented Tasks?
Do They Generate Their Training Corpus

Han^{1*}
Song Sun¹
Microsoft Research
Artificial Intelligence,
Technology,

Tsinghua University,
Department of
Human-Computer
Interaction and
User-Centered
Intelligence,
Beihang University,
Beijing, China
jiaohao@buaa.edu.cn

Gretchen Krueger Tom Henighan

Empirical Saturnalia



Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:
Algorithms for Recursive Self-Improvement through

DO LLMs HAVE

Naam
Tel.A

Sébastien B
Eric Horvitz

Self-Int
Internat

d.plu



IN THEY FOLLOW

oning Tasks

gence,

Oriented Tasks?

L Corso

n¹
elligence,
ogy,

iah*

h Sastry

Keerthi Reddy
Independent Researcher

Jorge Morales
Northeastern University

edu.cn

Gretchen Krueger Tom Henighan

Interpretability as a Natural Science

The Structure of Scientific Revolutions by Thomas Kuhn [42] is a classic text on the history and sociology of science. In it, Kuhn distinguishes between “normal science” in which a scientific community has a paradigm, and “extraordinary science” in which a community lacks a paradigm, either because it never had one or because it was weakened by crisis. It’s worth noting that “extraordinary science” is not a desirable state: it’s a period where researchers struggle to be productive.

Kuhn’s description of pre-paradigmatic fields feel eerily reminiscent of interpretability today.⁹ There isn’t consensus on what the objects of study are, what methods we should use to answer them, or how to evaluate research results. To quote a recent interview with Ian Goodfellow: “For interpretability, I don’t think we even have the right definitions.” [43]

One particularly challenging aspect of being in a pre-paradigmatic field is that there isn’t a shared sense of how to evaluate work in interpretability. There are two common proposals for dealing with this, drawing on the standards of adjacent fields. Some researchers, especially those with a deep learning background, want an “interpretability benchmark” which can evaluate how effective an interpretability method is. Other researchers with an HCI background may wish to evaluate interpretability methods through user studies.

But interpretability could also borrow from a third paradigm: natural science. In this view, neural networks are an object of empirical investigation, perhaps similar to an organism in biology. Such work would try to make empirical claims about a given network, which could be held to the standard of falsifiability.

(Olah et al., 2020)

The Empiricization of Computer Science

 MANOEL HORTA RIBEIRO
DEC 17, 2025

 15  6  3

Share

Yet the nature of computer science has changed tremendously: it is now an empirical science, driven by observation and experiment as much as by theory and construction. If you don’t believe me, spend 5 minutes going over the best paper awards for three conferences in different subfields in CS. If you did so in 2025, you might have found papers like “[Characterizing and Detecting Propaganda-Spreading Accounts on Telegram](#)” (USENIX; Security and Privacy), or “[Examining Mental Health Conversations with Large Language Models through Reddit Analysis](#)” (CSCW; Human Computer Interaction), or “[Scaling Depth Can Enable New Goal-Reaching Capabilities](#)” (NeuRIPS; Machine Learning). The first two papers describe online user traces, to understand sociotechnical phenomena (Online Propaganda, people using LLMs to talk about mental health issues), while the latter is a full-blown examination of what happens when you do self-supervised RL with very deep neural networks.

This empiricization of Computer Science seems to have happened independently across subdisciplines, and I argue that the reasons are two-fold.

(Horta Ribeiro, 2025)

Empiricist Turn in Computer Science

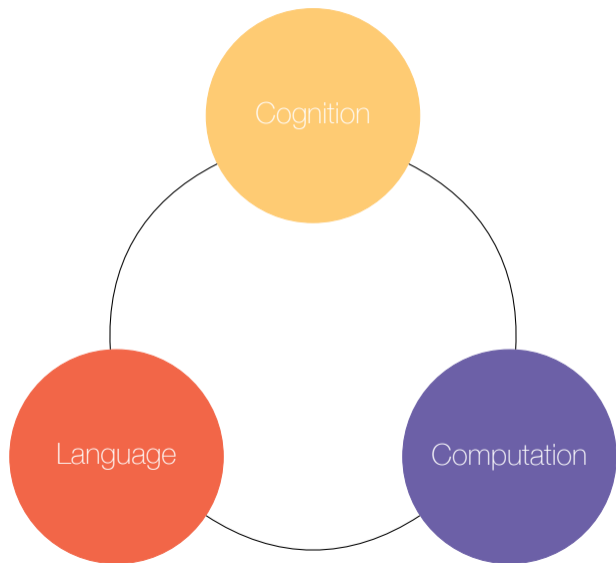
Chomsky's Trap

Language Models as Formal Objects

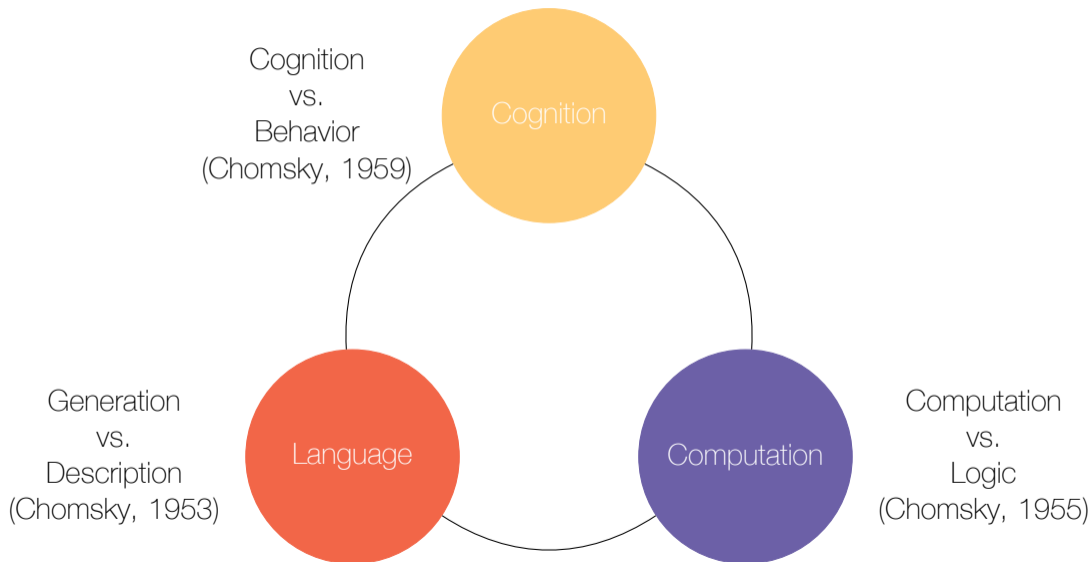
Philosophical Consequences

Takeaways

Chomsky's Generativist Program and the Cognitive Revolution



Chomsky's Generativist Program and the Cognitive Revolution

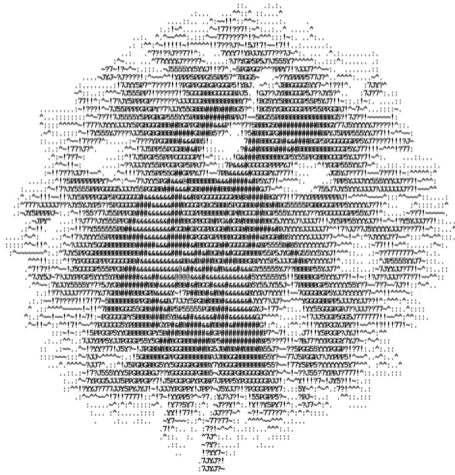


The New York Times

OPINION
GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023



Chomsky against Abstraction in Principle

“Pick the properties that you like for a set of processors. Pick the criteria you like for success, whether in terms of performance or structure or whatever. Consider the class of all organisms, *abstracting in principle* from the existing world, that satisfy those things. And then you can ask whether they have some property of things in the material world. Do they breathe? Do they grow? Do they think? Do they talk? Do they walk? Do they enjoy themselves? Do they have moral rights?”

(Chomsky, 1992)



Chomsky against Abstraction in Principle

“All of these questions are stupid. And the reason they’re stupid is because you’ve departed from naturalism. **Once you’ve departed from naturalism, you have an algorithm for constructing stupid questions.**”

(Chomsky, 1992)



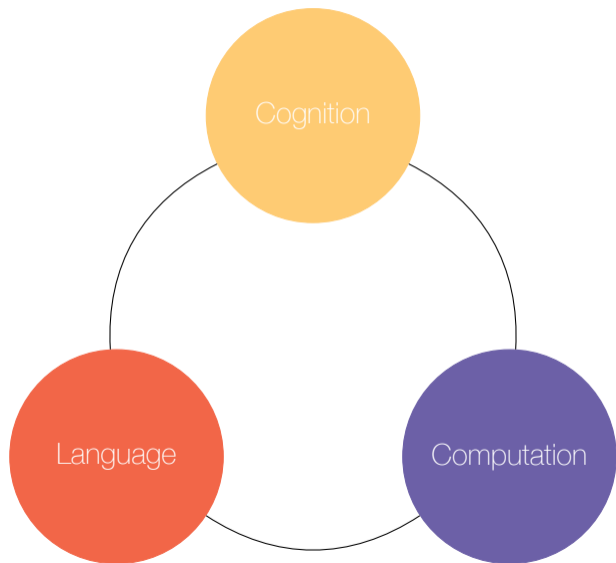
Chomsky against Abstraction in Principle

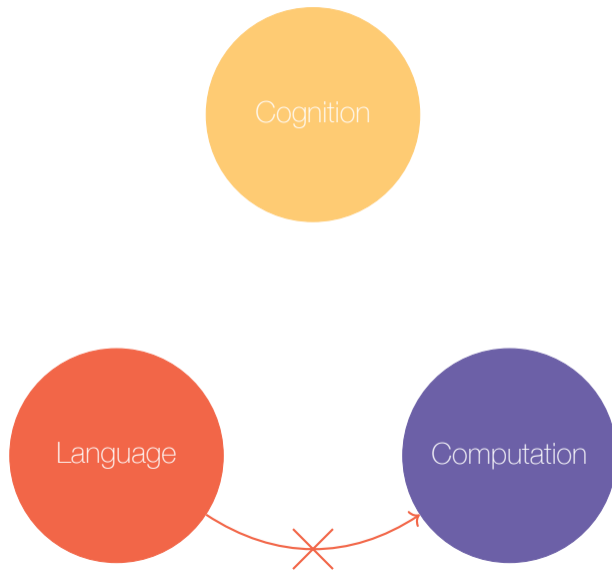
“There’s nothing wrong with principled abstraction. In fact, one might think of large areas of mathematics as that. But here we have something new, principled abstraction in an empirical discipline.”

“I don’t think we should cross that border, because there’s no empirical claim. It is just a question of how to extend the metaphor.”

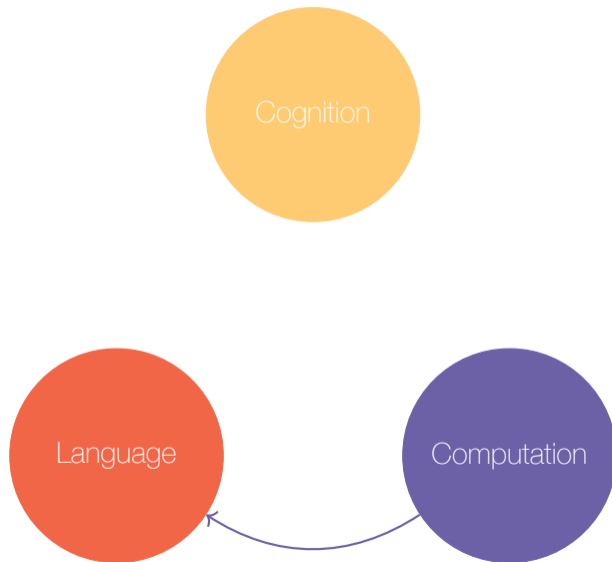
(Chomsky, 1992)



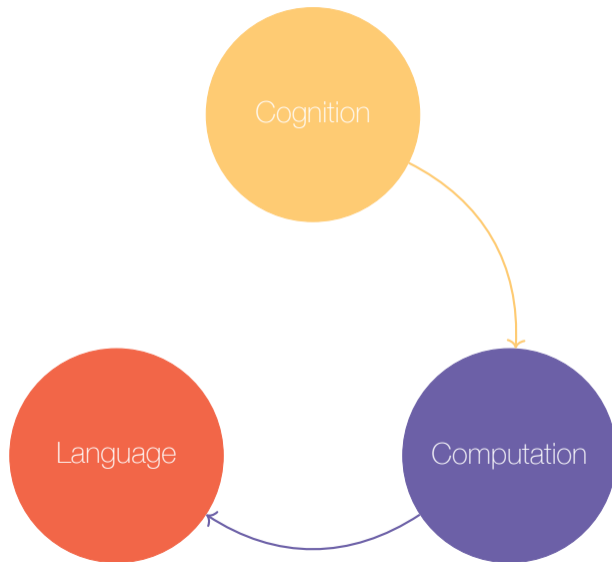


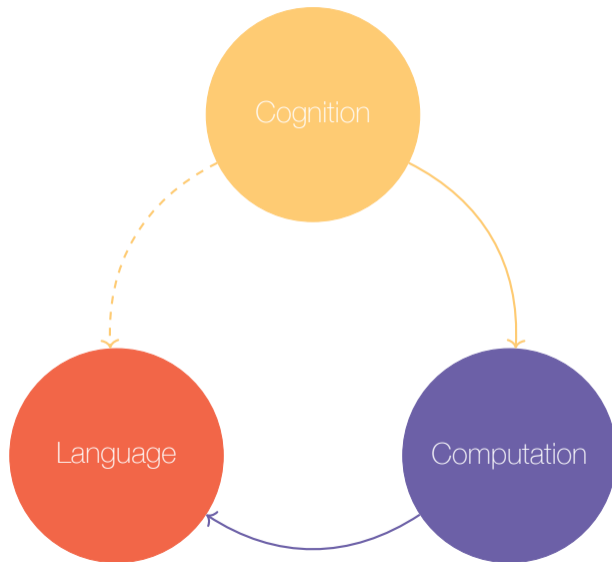


The Trap



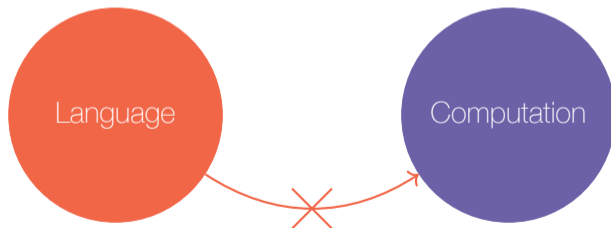
The Trap





Necessary Condition?

- ◇ Inadequacy of distributional models (Chomsky, 1953)
- ◇ Limited expressive power of FSAs (Chomsky, 1956)
- ◇ The probability of a sentence is useless (Chomsky, 1957, 1959)
- ◇ Poverty of stimulus (Chomsky, 1959)



Necessary Condition?

- ◇ Inadequacy of distributional models (Chomsky, 1953)

Inconclusive

- ◇ The probability of a sentence is useless (Chomsky, 1957, 1959)

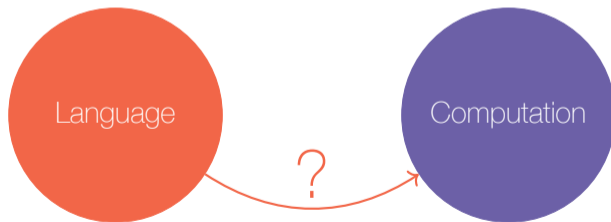
Empirically challenged

- ◇ Limited expressive power of FSAs (Chomsky, 1956)

The relevance is unclear

- ◇ Poverty of stimulus (Chomsky, 1959)

Assumes what is to be proved



Empiricist Turn in Computer Science

Chomsky's Trap

Language Models as Formal Objects

Philosophical Consequences

Takeaways

LLMs are computable functions

Neural LM



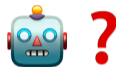
LLMs are computable functions

Neural LM



LLMs are computable functions

Neural LM



LLMs are computable functions

Neural LM



LLMs are computable functions

Neural LM



LLMs are computable functions

Neural LM



LLMs are computable functions

Neural LM



f !

LLMs are computable functions

Neural LM



Function

$$f: \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \dots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

LLMs are computable functions

Neural LM

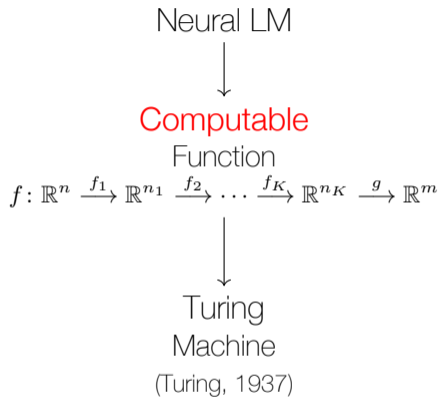


Computable

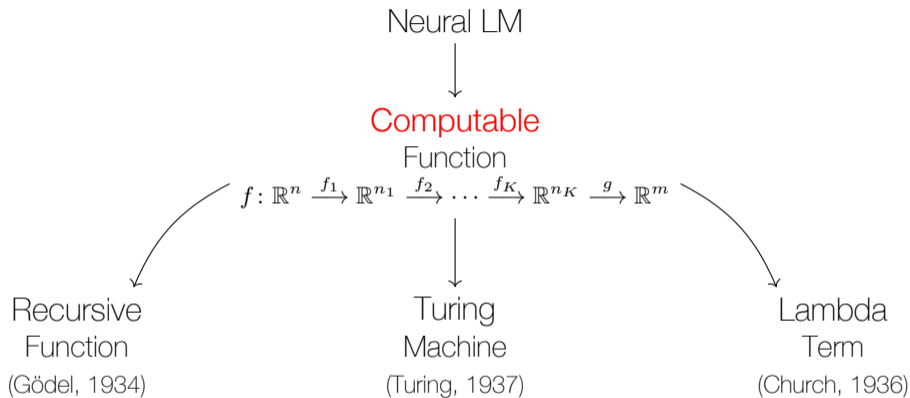
Function

$$f: \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \dots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

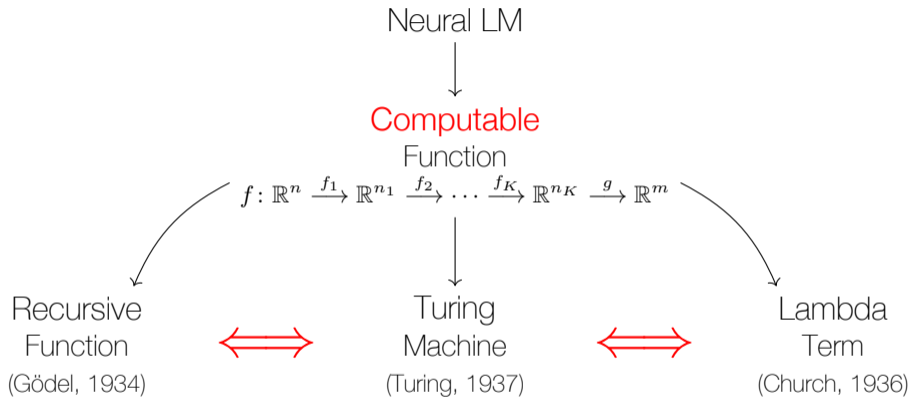
LLMs are computable functions



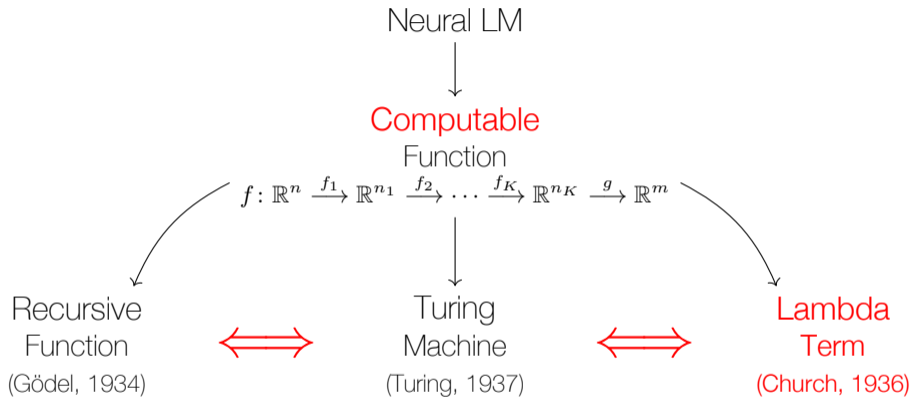
LLMs are computable functions



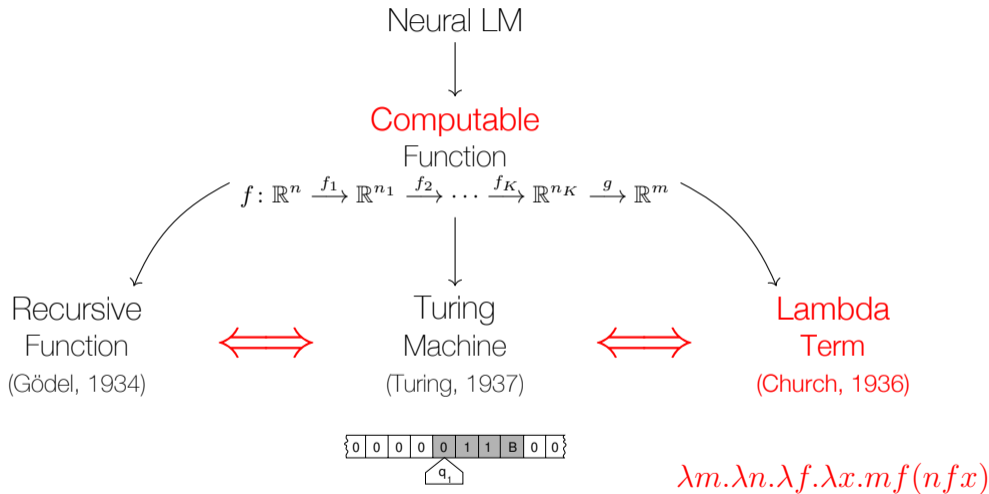
LLMs are computable functions



LLMs are computable functions



LLMs are computable functions



credit: Nynexman4464

$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$0: \lambda f. \lambda x. x$$

$$1: \lambda f. \lambda x. f x$$

$$2: \lambda f. \lambda x. f (f x)$$

$$3: \lambda f. \lambda x. f (f (f x))$$

$$4: \lambda f. \lambda x. f (f (f (f x)))$$

$$5: \lambda f. \lambda x. f (f (f (f (f x))))$$

...

$$n: \lambda f. \lambda x. \underbrace{f(\dots(f x)\dots)}_{n \text{ times}}$$

$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$

0: $\lambda f. \lambda x. x$

$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$

1: $\lambda f. \lambda x. f x$

2: $\lambda f. \lambda x. f (f x)$

3: $\lambda f. \lambda x. f (f (f x))$

4: $\lambda f. \lambda x. f (f (f (f x)))$

5: $\lambda f. \lambda x. f (f (f (f (f x))))$

...

$n: \lambda f. \lambda x. \underbrace{f(\dots(f x)\dots)}_{n \text{ times}}$

$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$

0: $\lambda f. \lambda x. x$

1: $\lambda f. \lambda x. f x$

2: $\lambda f. \lambda x. f (f x)$

3: $\lambda f. \lambda x. f (f (f x))$

4: $\lambda f. \lambda x. f (f (f (f x)))$

5: $\lambda f. \lambda x. f (f (f (f (f x))))$

...

$n: \lambda f. \lambda x. \underbrace{f(\dots(f x)\dots)}_{n \text{ times}}$

$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$

⋮

⋮

⋮

⋮

⋮

⋮

⋮

$\lambda f. \lambda x. f (f (f (f (f x))))$

$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$

$P' := \lambda r. \lambda s. \lambda f. \lambda x. f(f(f(f f x)))$

0: $\lambda f. \lambda x. x$

1: $\lambda f. \lambda x. f x$

2: $\lambda f. \lambda x. f(f x)$

3: $\lambda f. \lambda x. f(f(f x))$

4: $\lambda f. \lambda x. f(f(f(f x)))$

5: $\lambda f. \lambda x. f(f(f(f(f x))))$

...

$n: \lambda f. \lambda x. \underbrace{f(\dots(f x)\dots)}_{n \text{ times}}$

$\lambda r. \lambda s. \lambda f. \lambda x. f(f(f(f f x)))(\lambda f. \lambda x. f(f x))(\lambda f. \lambda x. f(f(f x)))$

↯

↯

↯

↯

↯

↯

↯

$\lambda f. \lambda x. f(f(f(f f x)))$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$0: \lambda f. \lambda x. x$$

$$1: \lambda f. \lambda x. f x$$

$$2: \lambda f. \lambda x. f (f x)$$

$$3: \lambda f. \lambda x. f (f (f x))$$

$$4: \lambda f. \lambda x. f (f (f (f x)))$$

$$5: \lambda f. \lambda x. f (f (f (f (f x))))$$

...

$$n: \lambda f. \lambda x. \underbrace{f(\dots(f x)\dots)}_{n \text{ times}}$$

$$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$$

⋮

⋮

⋮

⋮

⋮

⋮

⋮

$$\lambda f. \lambda x. f (f (f (f (f x))))$$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$0: \lambda f. \lambda x. x$$

$$1: \lambda f. \lambda x. f x$$

$$2: \lambda f. \lambda x. f (f x)$$

$$3: \lambda f. \lambda x. f (f (f x))$$

$$4: \lambda f. \lambda x. f (f (f (f x)))$$

$$5: \lambda f. \lambda x. f (f (f (f (f x))))$$

...

$$n: \lambda f. \lambda x. \underbrace{f(\dots(f x)\dots)}_{n \text{ times}}$$

$$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$$

$$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda g. \lambda y. g (g y)) (\lambda h. \lambda z. h (h (h z)))$$

$$\lambda n. \lambda f. \lambda x. (\lambda g. \lambda y. g (g y)) f (n f x) (\lambda h. \lambda z. h (h (h z)))$$

$$\lambda n. \lambda f. \lambda x. (\lambda g. \lambda y. g (g y)) f (n f x) (\lambda h. \lambda z. h (h (h z)))$$

$$\lambda f. \lambda x. (\lambda g. \lambda y. g (g y)) f ((\lambda h. \lambda z. h (h (h z)))) f x$$

$$\lambda f. \lambda x. (\lambda y. f (f y)) ((\lambda h. \lambda z. h (h (h z)))) f x$$

$$\lambda f. \lambda x. (\lambda y. f (f y)) ((\lambda z. f (f (f z)))) x$$

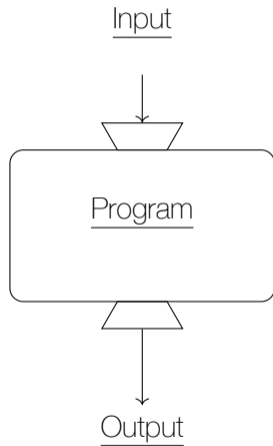
$$\lambda f. \lambda x. (\lambda y. f (f y)) (f (f (f x)))$$

$$\lambda f. \lambda x. f (f (f (f (f x))))$$

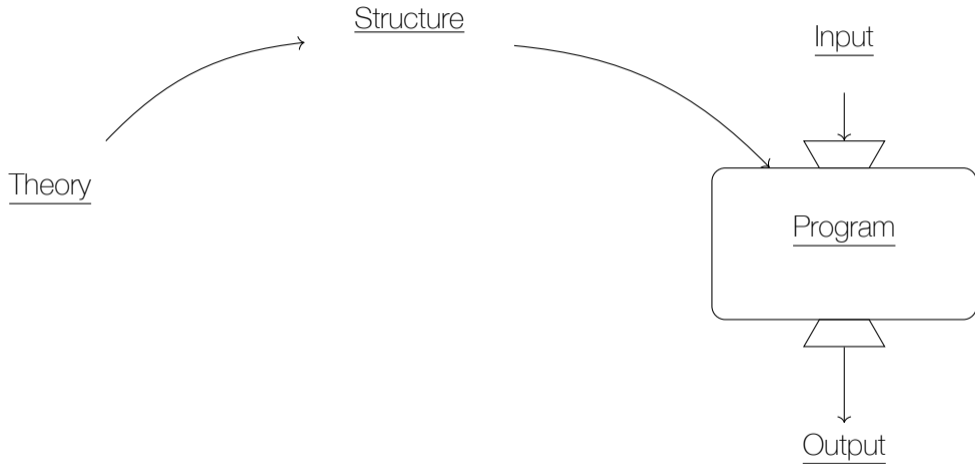
$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$

$P'' := \lambda R \acute{o} f \ddot{A} \ddot{O} \hat{e} \ddot{N} 5 \ddot{E} | \ddot{A} x \ddot{n} = \infty \ddot{u} \ddot{y} m W f 286 \ddot{e} y ' S \ddot{O} \acute{u} > v \& i \hat{A} - 2 \acute{o} \acute{E} 7 \acute{o} \zeta \infty \{ \ddot{a} > 2 f \ddot{B} \acute{o} \mu G \# \ddot{A} 9 \zeta U$
 $\infty \acute{o} b t Y \ddot{B} \hat{o} \ddot{Y} \ddot{U} \ddot{e} \% 0 3 ; 5 \acute{a} [l - \acute{e} u \hat{o} \ddot{U} \acute{e} \acute{7} - \ddot{U} . \lambda : \hat{^} 4 m \acute{O} \acute{O} \ddot{Y} ' \acute{e} - + \acute{I} s \acute{O} , \$ + g \ddot{i} , B^{TM} \div \acute{o} - \# i \ddot{Y} \hat{e} \hat{U} v$
 $- g \acute{O} \ddot{y} / \acute{e} i i j \acute{O} \ddot{f} \acute{C} e f i \bullet J 1 \ll \acute{E} \acute{o} , \acute{I} h \hat{e} t \ddot{f} \acute{a} e Y \$ \hat{^} 6 F i W \gg R \acute{U} K g e \ddot{r} . \lambda \ddot{f} d \ddot{r} \dots D 2 \div \acute{c} \acute{o} \acute{x} \acute{e} \acute{E} y . \acute{O} \ddot{r} c b$
 $B \acute{e} \acute{L} N \acute{E} 1 \hat{E} \ddot{f} / \hat{U} 9 \ddot{N} \mu - / J Y \zeta \acute{o} \acute{E} 9 \ddot{y} \hat{A} \acute{E} \acute{E} . \lambda \acute{A} \acute{I} \hat{A} \hat{^} \acute{o} \zeta , \gg f q \infty \pm \hat{i} \sim B 5 \hat{I} > \acute{O} \sim g^{TM} \acute{6} \Omega e \acute{a} \acute{e} C / \acute{a} \dots \acute{O}$
 $\cdot f \acute{O} \acute{A}] \ddot{N} \acute{a} y \hat{E} N \acute{E} \acute{e} \ddot{r} . \lambda \acute{A} \acute{e} \acute{a} \acute{e} f U \acute{o} \acute{f} E \acute{U} \acute{I} m \# , , 4 \backslash r \sqrt{-} \div \hat{I} p \acute{o} \gg y \ast v t \acute{A} J \acute{A} F 1 \hat{u} \acute{A} \acute{o} z \ll \acute{n} M \ddot{r} D j \acute{C} E$
 $B \acute{E} \acute{e} \acute{I} T _ \hat{E} a \% 0 \acute{A} \zeta \Omega @ \backslash \acute{O} \hat{^} \sim] \hat{I} \acute{h} \ddot{f} : \hat{^} 4 m \acute{O} \acute{O} \ddot{Y} ' \acute{e} - + \acute{I} s \acute{O} , \$ + g \ddot{i} , B^{TM} \div \acute{o} - \# i \ddot{Y} \hat{e} \hat{U} v - g \acute{O} \ddot{y}$
 $/ \acute{e} i i j \acute{O} \ddot{f} \acute{C} e f i \bullet J 1 \ll \acute{E} \acute{o} , \acute{I} h \hat{e} t \ddot{f} \acute{a} e Y \$ \hat{^} 6 F i W \gg R \acute{U} K g e \ddot{r} \acute{A} \acute{I} \hat{A} \hat{^} \acute{o} \zeta , \gg f q \infty \pm \hat{i} \sim B 5 \hat{I} > \acute{O} \sim g^{TM} \acute{6}$
 $\Omega e \acute{a} \acute{e} C / \acute{a} \dots \acute{O} \cdot f \acute{O} \acute{A}] \ddot{N} \acute{a} y \hat{E} N \acute{E} \acute{e} \ddot{r} (\ddot{f} d \ddot{r} \dots D 2 \div \acute{c} \acute{o} \acute{x} \acute{e} \acute{E} y . \acute{O} \ddot{r} c b B \acute{e} \acute{L} N \acute{E} 1 \hat{E} \ddot{f} / \hat{U} 9 \ddot{N} \mu - /$
 $J Y \zeta \acute{o} \acute{E} 9 \ddot{y} \hat{A} \acute{E} \acute{E} \acute{A} \acute{I} \hat{A} \hat{^} \acute{o} \zeta , \gg f q \infty \pm \hat{i} \sim B 5 \hat{I} > \acute{O} \sim g^{TM} \acute{6} \Omega e \acute{a} \acute{e} C / \acute{a} \dots \acute{O} \cdot f \acute{O} \acute{A}] \ddot{N} \acute{a} y \hat{E} N \acute{E} \acute{e} \ddot{r} \acute{A}$
 $\acute{e} \acute{f} U \acute{o} \acute{f} E \acute{U} \acute{I} m \# , , 4 \backslash r \sqrt{-} \div \hat{I} p \acute{o} \gg y \ast v t \acute{A} J \acute{A} F 1 \hat{u} \acute{A} \acute{o} z \ll \acute{n} M \ddot{r} D j \acute{C} E B \acute{E} \acute{e} \acute{I} T _ \hat{E} a \% 0 \acute{A} \zeta \Omega @ \backslash$
 $\acute{O} \hat{^} \sim] \hat{I} \acute{h} \ddot{f}) (\acute{E} \hat{I} \hat{U} \acute{e} i 4 W \mu \acute{I} \} w , , \$ \Omega \acute{K} 5 \acute{e} \acute{A} \acute{Q} \% 3 [m \acute{r} \sim B \acute{A} f i \acute{O} ; \acute{o} J \zeta \acute{C} \acute{E} \hat{i} \acute{o} \ddot{Y} \acute{O} c B , \acute{n} \$ \acute{A} \acute{a} \} \acute{O} \acute{A} \acute{O} 3 ;$
 $\acute{r} ? \acute{o} \acute{o} \acute{C} E @ f \acute{l} 8 \acute{r} R \acute{C} \acute{A} \acute{o} \acute{r} \ast \& < \acute{Y} - \acute{o} 1 2 \acute{A} \% 0 \acute{a} \acute{O} \acute{U} \# \acute{i} \acute{r} , \acute{u} \acute{r} \ll \acute{o} , , \infty \acute{I} \acute{a} \acute{a} \acute{e} \acute{o} \acute{A} d | \acute{r} \acute{N} \acute{r} \acute{E} y \acute{O} ; \hat{^} W$
 $\acute{r} \acute{w} \acute{o} [] \backslash \gg \acute{O} \acute{E} \acute{u} w \acute{r} 6 < \acute{u} \acute{r} = \acute{a} \acute{O} \acute{r} \acute{I} \acute{D} z ? 2 \pm | \acute{e} \acute{r} \acute{3} \hat{A} / r x \mu \infty \mu \$ \acute{A} \acute{e} \hat{A} \acute{r} \acute{f} \acute{r} \hat{u} \acute{r} + \acute{I} V \acute{i} y \acute{a} G \acute{a} \acute{e} \acute{B} \acute{a} g \acute{o} / , u \acute{N}$

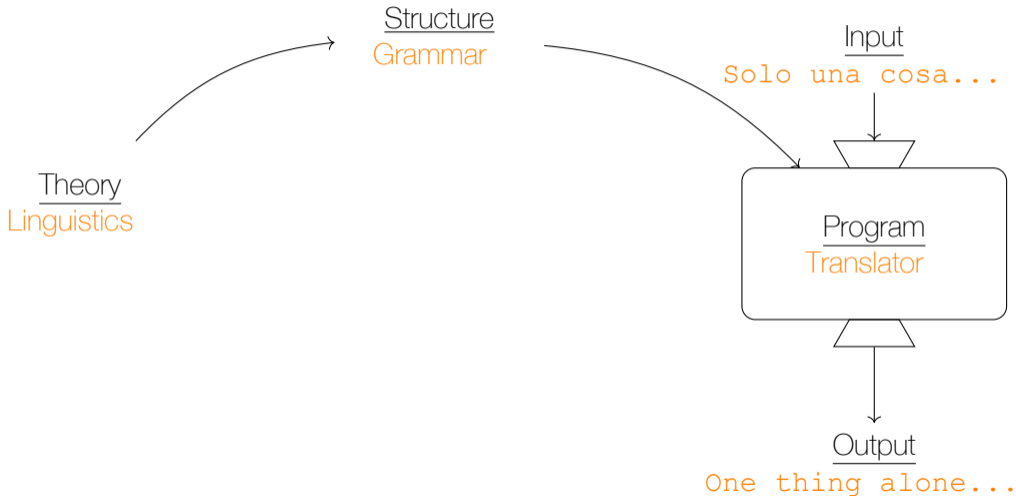
Implicit Structure



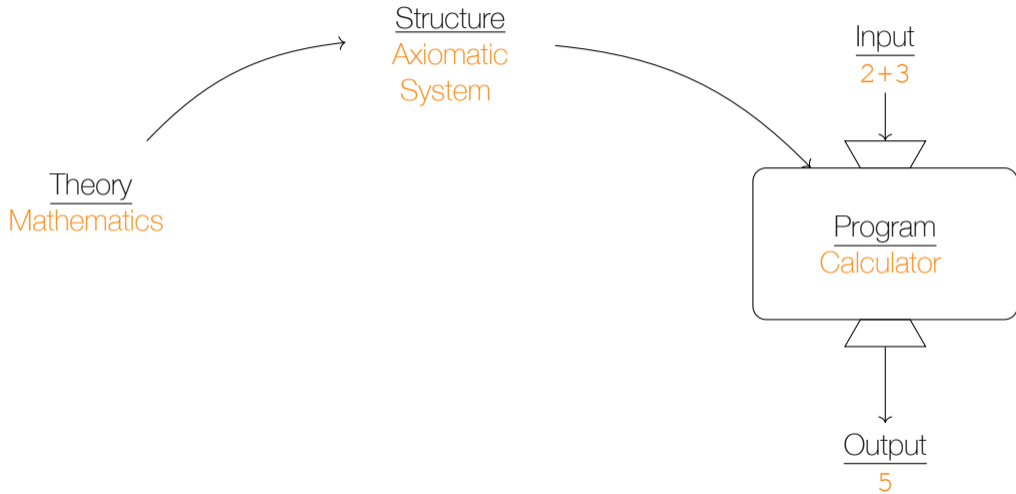
Implicit Structure



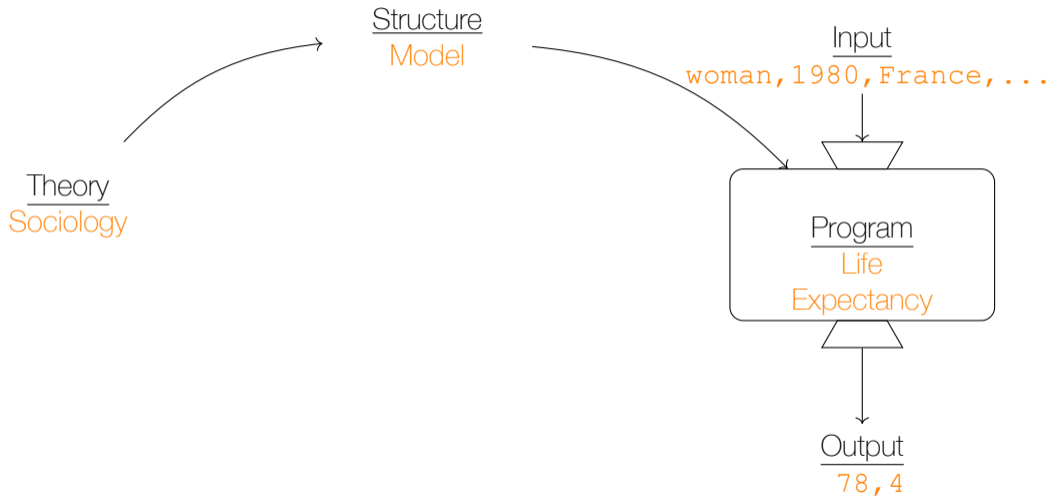
Implicit Structure



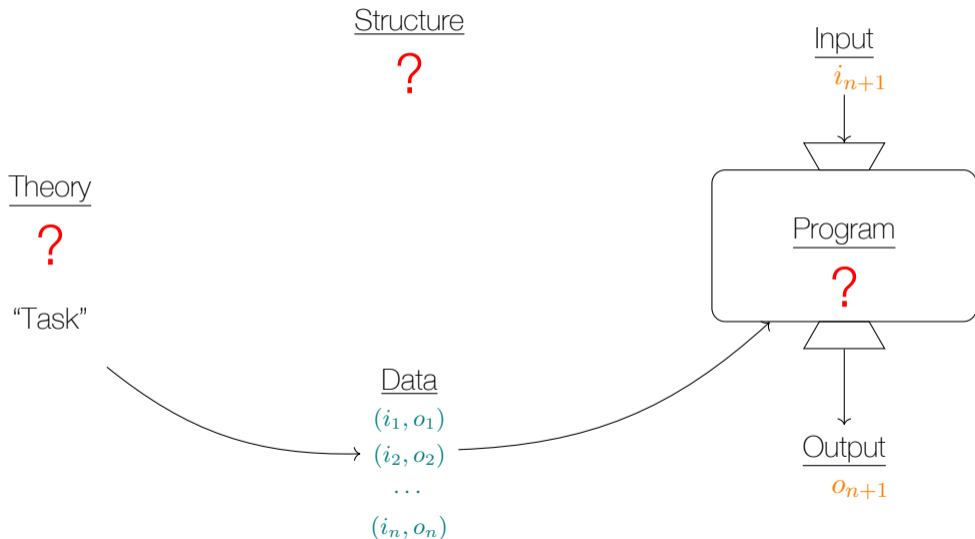
Implicit Structure



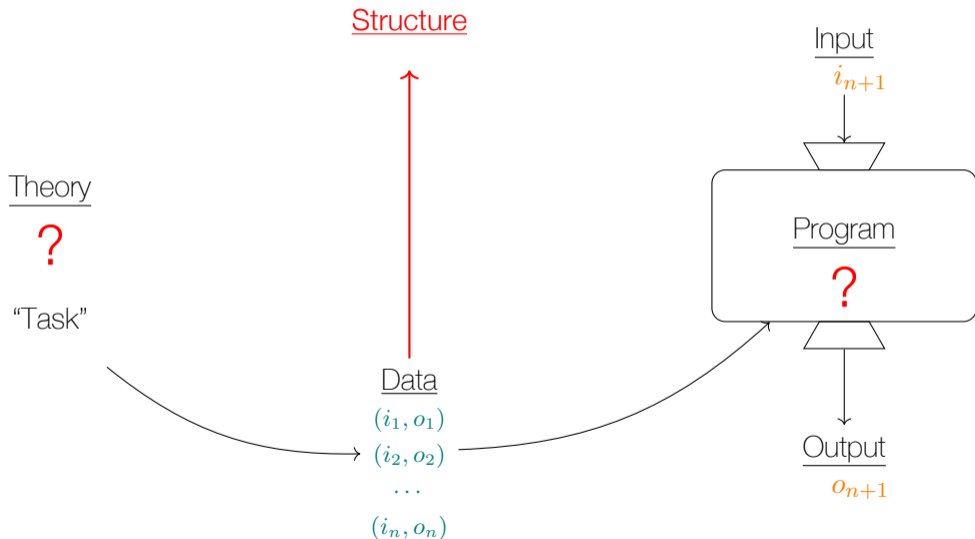
Implicit Structure



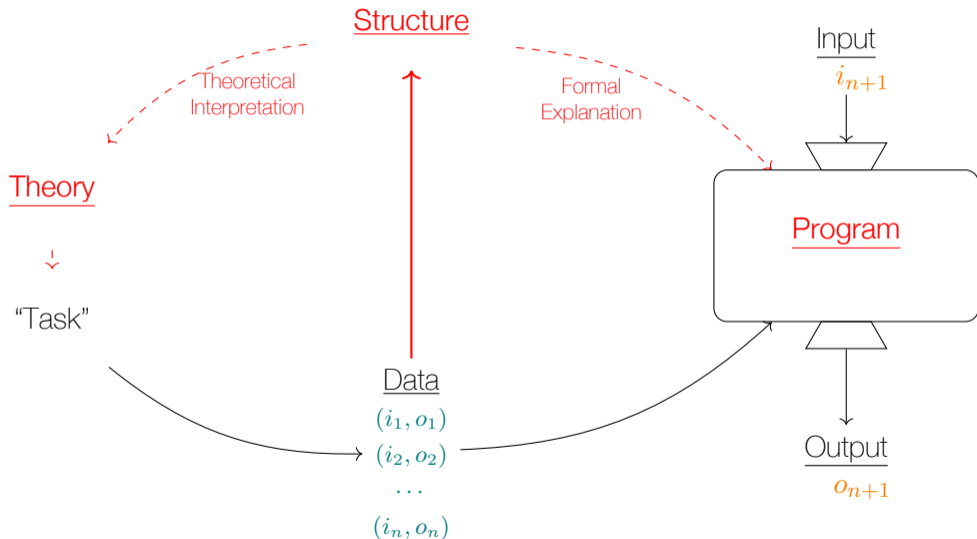
Implicit Structure



Making It Explicit



Making It Explicit

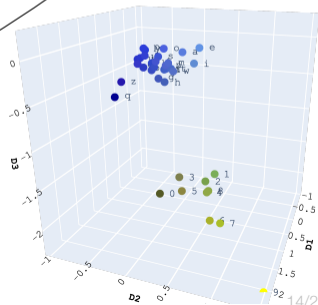
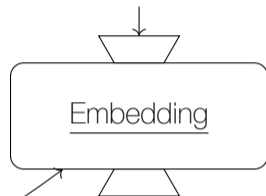


Embedding structure

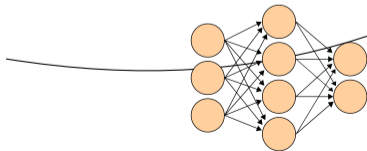
Structure

?

$\{-, /, 0, 1, 2, \dots, 8, 9, =,$
 $a, b, c, \dots, w, x, y, z, \acute{e}\}$



Data

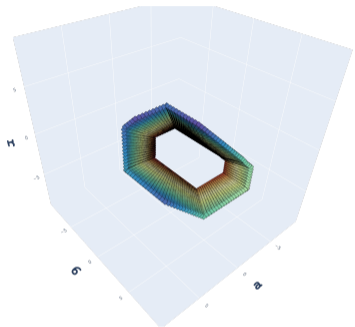


Structure

?

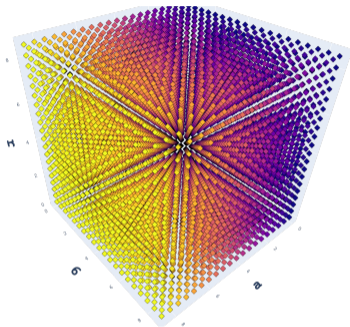
$$\begin{array}{ccc}
 \mathbf{C}^{\text{op}} \times \mathbf{D} & \rightarrow & \bar{\mathbb{R}} \\
 & \Downarrow & \\
 \mathcal{M}^* : \bar{\mathbb{R}}^{\mathbf{C}^{\text{op}}} & \rightleftarrows & (\bar{\mathbb{R}}^{\mathbf{D}})^{\text{op}} : \mathcal{M}_*
 \end{array}$$

Structure



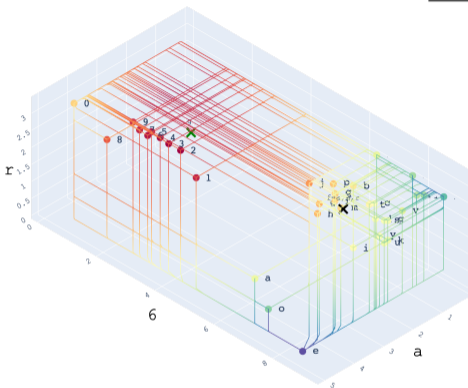
?

$\mathcal{M}_* \mathcal{M}^*$



$$\begin{array}{ccc}
 \mathbb{C}^{\text{op}} \times \mathbb{D} & \rightarrow & \bar{\mathbb{R}} \\
 & \rightleftharpoons & \\
 \mathcal{M}^* : \bar{\mathbb{R}}^{\mathbb{C}^{\text{op}}} & \rightleftharpoons & (\bar{\mathbb{R}}^{\mathbb{D}})^{\text{op}} : \mathcal{M}_*
 \end{array}$$

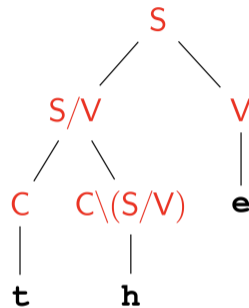
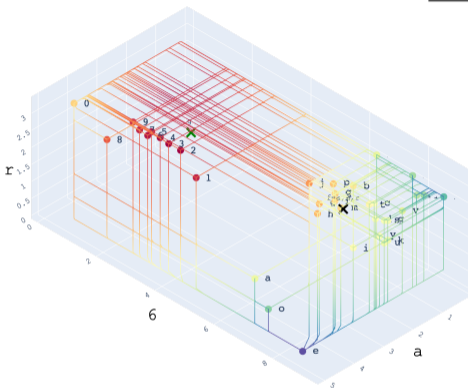
Structure



$$\begin{array}{ccc}
 \mathbb{C}^{\text{op}} \times \mathbb{D} & \rightarrow & \bar{\mathbb{R}} \\
 & \Downarrow & \\
 \mathcal{M}^* : \bar{\mathbb{R}}^{\mathbb{C}^{\text{op}}} & \Longleftrightarrow & (\bar{\mathbb{R}}^{\mathbb{D}})^{\text{op}} : \mathcal{M}_*
 \end{array}$$

A red arrow points from the bottom-left node of the diagram above to the $\bar{\mathbb{R}}^{\mathbb{C}^{\text{op}}}$ term in the equation.

Structure



$$\begin{array}{c}
 \mathbb{C}^{\text{op}} \times \mathbb{D} \rightarrow \bar{\mathbb{R}} \\
 \Downarrow \\
 \mathcal{M}^* : \bar{\mathbb{R}}^{\mathbb{C}^{\text{op}}} \iff (\bar{\mathbb{R}}^{\mathbb{D}})^{\text{op}} : \mathcal{M}_*
 \end{array}$$

Empiricist Turn in Computer Science

Chomsky's Trap

Language Models as Formal Objects

Philosophical Consequences

Takeaways

Stochastic parrots vs. AGI



LLMs are not like us,
therefore they do not and can not have any relation to natural language.



LLMs have a relation to natural language,
therefore they are like us.

...I supposed that all the objects (presentations) that had ever entered into my mind when awake, had in them no more truth than the illusions of my dreams. But immediately upon this I observed that, whilst I thus wished to think that all was false, it was absolutely necessary that I, who thus thought, should be something; And as I observed that this truth, I think, therefore I am, was so certain and of such evidence that no ground of doubt, however extravagant, could be alleged by the Sceptics capable of shaking it, I concluded that I might, without scruple, accept it as the first principle of the philosophy of which I was in search.

Descartes, *Meditations on First Philosophy* (1641)

But I was persuaded that there was nothing in all the world, that there was no heaven, no earth, that there were no minds, nor any bodies: was I not then likewise persuaded that I did not exist? Not at all; of a surety I myself did exist since I persuaded myself of something [or merely because I thought of something]. But there is some deceiver or other, very powerful and very cunning, who ever employs his ingenuity in deceiving me. Then without doubt I exist also if he deceives me, and let him deceive me as much as he will, he can never cause me to be nothing so long as I think that I am something. So that after having reflected well and carefully examined all things, we must come to the definite conclusion that this proposition: I am, I exist, is necessarily true each time that I pronounce it, or that I mentally conceive it.

Descartes, *Meditations on First Philosophy* (1641)

Language vs. thought

...the philosopher has to say: “When I dissect the process expressed in the proposition ‘I think,’ I get a whole set of bold claims that are difficult, perhaps impossible, to establish, – for instance, that I am the one who is thinking, that there must be something that is thinking in the first place, that thinking is an activity and the effect of a being who is considered the cause, that there is an ‘I,’ and finally, that it has already been determined what is meant by thinking, – that I know what thinking is. [...]

Nietzsche, *Beyond Good and Evil*, §16 (1886)

Language vs. thought

...Because if I had not already made up my mind what thinking is, how could I tell whether what had just happened was not perhaps 'willing' or 'feeling'? Enough: this 'I think' presupposes that I compare my present state with other states that I have seen in myself, in order to determine what it is: and because of this retrospective comparison with other types of 'knowing,' this present state has absolutely no 'immediate certainty' for me." – In place of that "immediate certainty" which may, in this case, win the faith of the people, the philosopher gets handed a whole assortment of metaphysical questions, genuinely probing intellectual questions of conscience, such as: "Where do I get the concept of thinking from? Why do I believe in causes and effects? What gives me the right to speak about an I, and, for that matter, about an I as cause, and, finally, about an I as the cause of thoughts?" [...]

Nietzsche, *Beyond Good and Evil*, §16 (1886)

Language vs. thought

Now in order to cognize ourselves, there is required in addition to the act of thought, which brings the manifold of every possible intuition to the unity of apperception, a determinate mode of intuition, whereby this manifold is given; it therefore follows that although my existence is not indeed appearance (still less mere illusion), the determination of my existence can take place only in conformity with the form of inner sense, according to the special mode in which the manifold, which I combine, is given in inner intuition. Accordingly I have no cognition of myself as I am but merely as I appear to myself

Kant, *Critique of Pure Reason* (1781)

Language vs. thought

But, isn't thinking a kind of speaking? How is it possible for thinking to be engaged in a struggle with speaking? Wouldn't that be a struggle in which thinking was at war with itself? Doesn't this spell the end to the possibility of thinking?

Frege, *Sources of Knowledge of Math. and the math. natural Sc.* (1924-25)

Language vs. thought

It is sometimes said: animals do not talk because they lack the mental abilities. And this means: “They do not think, and that is why they do not talk.” But — they simply do not talk.

Wittgenstein, *Philosophical Investigations*, 1953, § 25

Language vs. thought

The perennial man in the street believes that when he speaks he freely puts together whatever elements have the meanings he intends; but he does so only by choosing members of those classes that regularly occur together, and in the order in which these classes occur. [...] the restricted distribution of classes persists for all their occurrences; the restrictions are not disregarded arbitrarily, e.g. for semantic needs.

Harris, *Distributional Structure*, pp. 775-776, (1954).

Formal Content

(Gastaldi and Pellissier, 2021)

Form ~~vs.~~ ~~and~~ ~~Meaning~~ ~~Content~~

Kant, Hegel, Frege, Russian formalists, Saussure, Hjelmslev, etc.

Formal Content: The dimension of content which finds its source in the internal relations holding between the expressions of a language.

Formal Content

(Gastaldi and Pellissier, 2021)

Form ~~vs.~~ ~~and~~ ~~Meaning~~ Content

Kant, Hegel, Frege, Russian formalists, Saussure, Hjeltslev, etc.

Formal Content: The dimension of content which finds its source in the internal relations holding between the expressions of a language.

- ◇ Characteristic Content: The content resulting from the **inclusion** of a unit **in a class of other units** by which it accepts to be substituted in given contexts
- ◇ Syntactic Content: The content a unit receives as a result of the multiple **dependencies** it can maintain with respect **to other units** in its context
- ◇ Informational Content: The content related to the **non-uniform distribution of units** within those substitutability classes

Illustration of Formal Contents

Characteristic Content

```
{cat, dog, spider,  
 gavagai}
```

Atomic Type

Syntactic Content

```
"the gavagai is on the  
   mat"
```

Profunctor Nucleus

Informational Content

```
{cat:0.059%,  
 dog:0.012%,  
 spider:0.009%,  
 gavagai:0.000%}
```

Probability Distribution

(Gastaldi & Pellissier, 2021)

Empiricist Turn in Computer Science

Chomsky's Trap

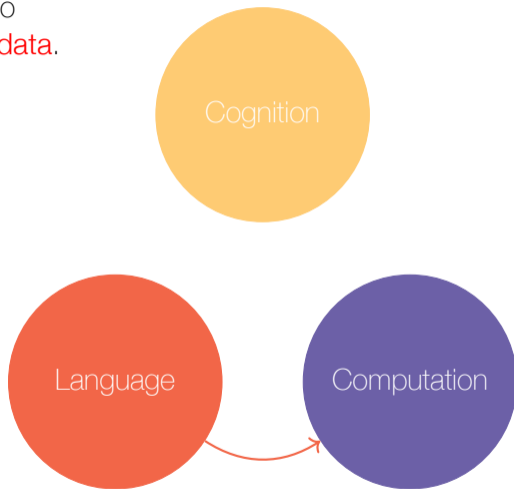
Language Models as Formal Objects

Philosophical Consequences

Takeaways

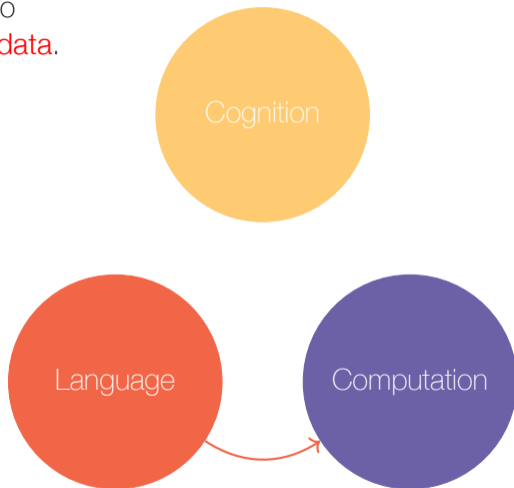
Takeaways

- ◊ A **formal** approach to data analysis can contribute to inferring **symbolic language** models **from** linguistic **data**.



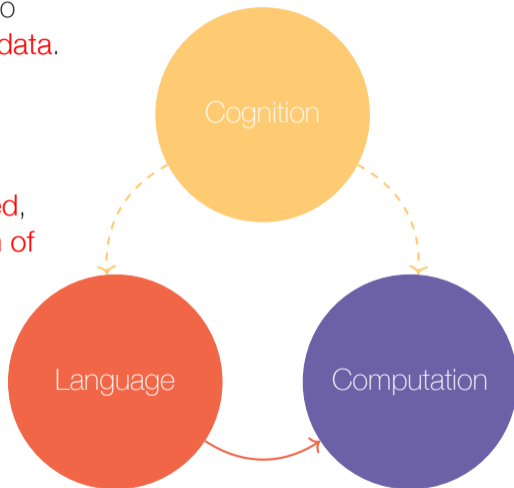
Takeaways

- ◇ A **formal** approach to data analysis can contribute to inferring **symbolic language** models **from** linguistic **data**.
- ◇ Resulting models are, a priori, **models of the data**.



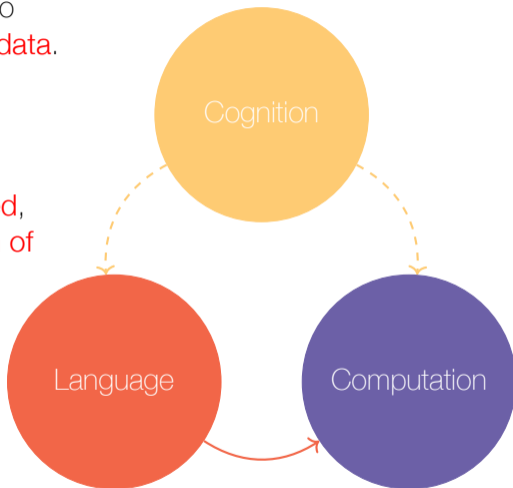
Takeaways

- ◇ A **formal** approach to data analysis can contribute to inferring **symbolic language** models **from** linguistic **data**.
- ◇ Resulting models are, a priori, **models of the data**.
- ◇ The **cognitive content** of such models is **suspended**, and cannot be restored without raising the **problem of the data**.



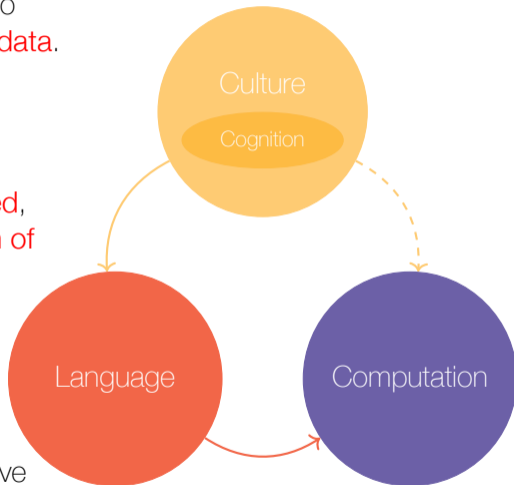
Takeaways

- ◇ A **formal** approach to data analysis can contribute to inferring **symbolic language** models **from** linguistic **data**.
- ◇ Resulting models are, a priori, **models of the data**.
- ◇ The **cognitive content** of such models is **suspended**, and cannot be restored without raising the **problem of the data**.
- ◇ The **scale** of the data for such models **exceeds the individual scale**.



Takeaways

- ◇ A **formal** approach to data analysis can contribute to inferring **symbolic language** models **from** linguistic **data**.
- ◇ Resulting models are, a priori, **models of the data**.
- ◇ The **cognitive content** of such models is **suspended**, and cannot be restored without raising the **problem of the data**.
- ◇ The **scale** of the data for such models **exceeds the individual scale**.
- ◇ **Cultural conditions** of data production become **constitutive** in the relation between cognitive contents and language models.



References I

- Chomsky, N. (1953). Systems of syntactic analysis. *Journal of Symbolic Logic*, 18(3), 242–256.
<https://doi.org/10.2307/2267409>
- Chomsky, N. (1955). Logical syntax and semantics: Their linguistic relevance. *Language*, 31(1), 36–45.
- Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2(3), 113–124. <https://doi.org/10.1109/TIT.1956.1056813>
- Chomsky, N. (1957). *Syntactic structures*. Mouton; Co.
- Chomsky, N. (1959). *Language*, 35(1), 26–58. Retrieved July 7, 2025, from <http://www.jstor.org/stable/411334>
- Chomsky, N. (1992, November). Language and the “cognitive revolutions” [Delivered November 23–27, 1992].
- Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2), 345–363.
- Gastaldi, J. L., & Pellissier, L. (2021). The calculus of language: explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*, 46(4), 569–590.
<https://doi.org/10.1080/03080188.2021.1890484>
- Gödel, K. (1934). On undecidable propositions of formal mathematical systems. In *Collected works* (pp. 346–371). Clarendon Press Oxford University Press.
- Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020). Zoom in: An introduction to circuits [<https://distill.pub/2020/circuits/zoom-in>]. *Distill*. <https://doi.org/10.23915/distill.00024.001>
- Turing, A. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1), 230–265. <https://doi.org/10.1112/plms/s2-42.1.230>
- Wittgenstein, L. (1953). *Philosophical investigations* (4th ed.). Wiley-Blackwell.

CNRS French–Danish Workshop
Reasoning in the Embedding Space
Copenhagen, Denmark

Empiricism vs. Formalism

In the Study of Embeddings

Juan Luis Gastaldi

`www.giannigastaldi.com`

ETH zürich

January 20, 2026