

ITS, CUNY Graduate Center  
New York, NY, USA

*Toward a Critical Formalism*  
From the Philosophy to the Theory  
of Distributional Language Models

Juan Luis Gastaldi

**ETH** zürich

April 11, 2025

# Outline

Introduction

NLMs as Formal Objects

Conclusion

# Outline

Introduction

NLMs as Formal Objects

Conclusion

# The Epistemology of AI

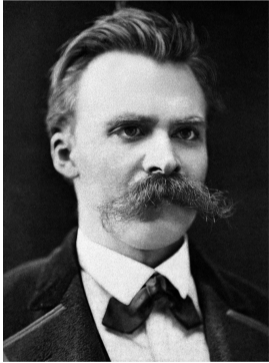


# Where Art Thou, Critique?

# Where Art Thou, Critique?

- ◇ Good “externalist” critique
- ◇ Poor “internalist” critique
  - The main “critical” reference remains the “Stochastic Parrots” approach (Bender & Koller, 2020; Bender et al., 2021)
  - Kirschenbaum (2023):  
Bender et al.’s (2021) paper “offers a disarming linear account of how language, communication, intention, and meaning work, one that would seem to sidestep decades of scholarship around these same issues in literary theory [...] the passage would be red meat for a graduate critical-theory seminar.”
  - Underwood (2023):  
“The beautiful irony of this situation [...] is that a generation of humanists trained on Foucault have now rallied around “On the Dangers of Stochastic Parrots” to oppose a theory of language that their own disciplines invented, just at the moment when computer scientists are reluctantly beginning to accept it.”

# The Birth of Contemporary Critique

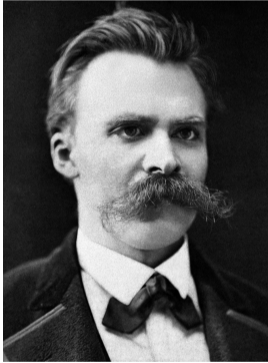


“In some remote corner of the universe, flickering in the light of the countless solar systems into which it had been poured, there was once a planet on which **clever animals invented cognition**. It was the most **arrogant** and most **mendacious** minute in the ‘history of the world’...”

“On Truth and Lying in a Non-Moral Sense”  
(Nietzsche, 1873)

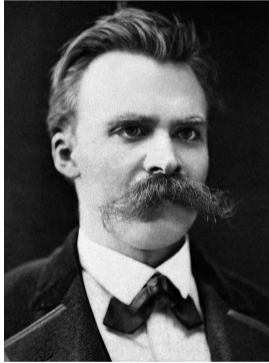
# The Birth of Contemporary Critique

- ◇ “The ‘thing-in-itself’ ([...] truth without consequences) is impossible for even the creator of language to grasp”



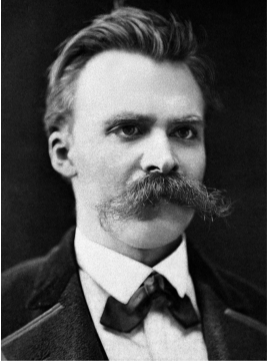


# The Birth of Contemporary Critique



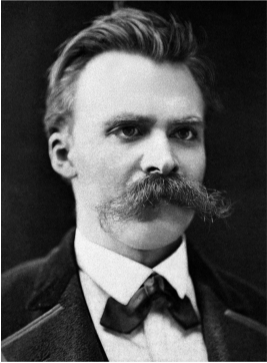
- ◇ “The ‘thing-in-itself’ ([...] truth without consequences) is impossible for even the creator of language to grasp”
- ◇ “The stimulation of a nerve is first translated into an image: first metaphor!”

# The Birth of Contemporary Critique



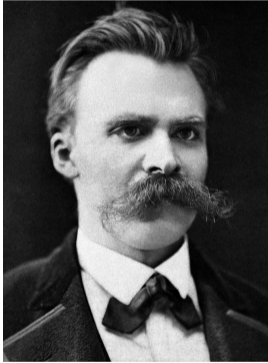
- ◇ “The ‘thing-in-itself’ ([...] truth without consequences) is impossible for even the creator of language to grasp”
- ◇ “The stimulation of a nerve is first translated into an image: first metaphor!”
- ◇ “The image is then imitated by a sound: second metaphor!”

# The Birth of Contemporary Critique



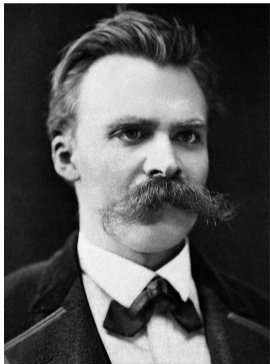
- ◇ “The ‘thing-in-itself’ ([...] truth without consequences) is impossible for even the creator of language to grasp”
- ◇ “The stimulation of a nerve is first translated into an image: first metaphor!”
- ◇ “The image is then imitated by a sound: second metaphor!”
- ◇ “...how arbitrary these translations are!”; “...inherently not a necessary relationship...”

# The Birth of Contemporary Critique



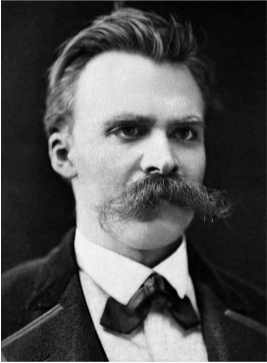
- ◇ “The ‘thing-in-itself’ ([...] truth without consequences) is impossible for even the creator of language to grasp”
- ◇ “The stimulation of a nerve is first translated into an image: first metaphor!”
- ◇ “The image is then imitated by a sound: second metaphor!”
- ◇ “...how arbitrary these translations are!”; “...inherently not a necessary relationship...”
- ◇ “What is a word? The copy of a nervous stimulation in sounds.”

# The Birth of Contemporary Critique



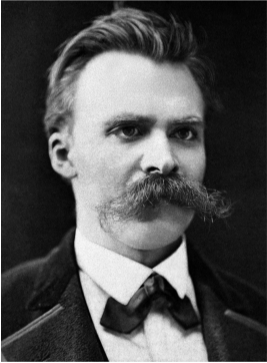
- ◇ “The ‘thing-in-itself’ ([...] **truth** without consequences) is **impossible** for even the creator of **language** to grasp”
- ◇ “The **stimulation** of a nerve is first translated into an **image**: first metaphor!”
- ◇ “The **image** is then imitated by a **sound**: second metaphor!”
- ◇ “...how **arbitrary** these translations are!”; “...inherently **not** a **necessary** relationship...”
- ◇ “What is a **word**? The copy of a nervous stimulation in **sounds**.”
- ◇ “...each **word** immediately becomes a **concept** [...] because at the same time it must fit [...] nothing other than nonequivalent cases.”

# The Birth of Contemporary Critique



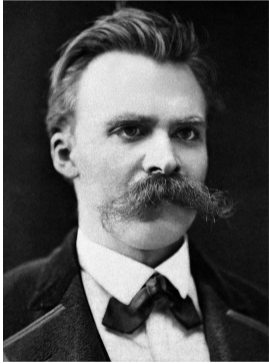
- ◇ “What, then, is **truth**? [...] a sum of **human relations**...”

# The Birth of Contemporary Critique



- ◇ “What, then, is **truth**? [...] a sum of **human relations**...”
- ◇ “...because necessity and boredom also lead men to want to live in **societies and herds**, they need a **peace treaty**...”

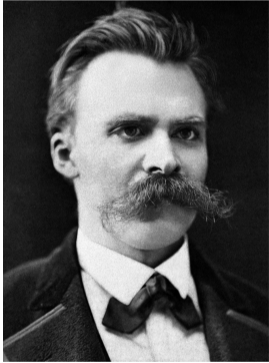
# The Birth of Contemporary Critique



- ◇ “What, then, is **truth**? [...] a sum of **human relations**...”
- ◇ “...because necessity and boredom also lead men to want to live in **societies and herds**, they need a **peace treaty**...”
- ◇ “...**society** imposes [...] the obligation to **lie in accordance** with firmly established **convention**...”

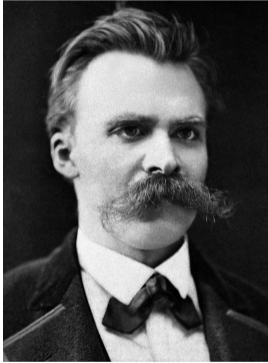


# The Birth of Contemporary Critique



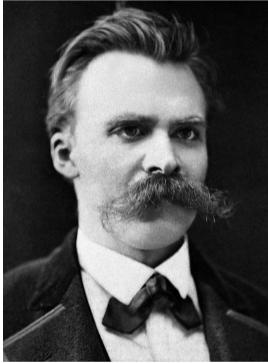
- ◇ “What, then, is **truth**? [...] a sum of **human relations**...”
- ◇ “...because necessity and boredom also lead men to want to live in **societies and herds**, they need a **peace treaty**...”
- ◇ “...**society** imposes [...] the obligation to **lie in accordance** with firmly established **convention**...”
- ◇ “...truths are **illusions** of which we have **forgotten** that they are illusions...”

# The Birth of Contemporary Critique



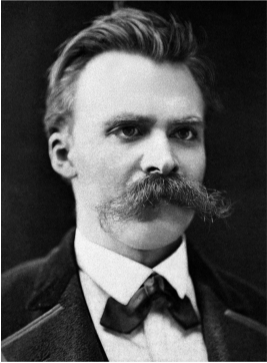
- ◇ “What, then, is **truth**? [...] a sum of **human relations**...”
- ◇ “...because necessity and boredom also lead men to want to live in **societies and herds**, they need a **peace treaty**...”
- ◇ “...**society** imposes [...] the obligation to **lie in accordance** with firmly established **convention**...”
- ◇ “...truths are **illusions** of which we have **forgotten** that they are illusions...”
- ◇ “...human beings **forget** that this is how things are; thus they lie **unconsciously**...”

# The Birth of Contemporary Critique



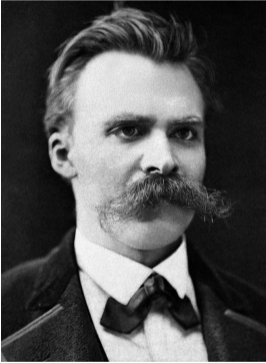
- ◇ “What, then, is **truth**? [...] a sum of **human relations**...”
- ◇ “...because necessity and boredom also lead men to want to live in **societies and herds**, they need a **peace treaty**...”
- ◇ “...**society** imposes [...] the obligation to **lie in accordance** with firmly established **convention**...”
- ◇ “...truths are **illusions** of which we have **forgotten** that they are illusions...”
- ◇ “...human beings **forget** that this is how things are; thus they lie **unconsciously**...”
- ◇ “...precisely *because of this* **unconsciousness**, precisely because of this forgetting, they arrive at the **feeling of truth**.”

# The Birth of Contemporary Critique



- ◇ “between [...] subject and object [...] there is [...] at most an *aesthetic* way of relating”

# The Birth of Contemporary Critique



- ◇ “between [...] subject and object [...] there is [...] at most an *aesthetic* way of relating”
- ◇ “Where the man of intuition [...] wields his weapons more mightily and victoriously than his contrary, a **culture** can take shape [...] and the rule of **art over life** can become established”

# Criticism and Formalism

- ◇ The argumentative matrix assembled by Nietzsche...
  - ...represented a radical **renewal of the 19th century critical project** in Western thought

# Criticism and Formalism

- ◇ The argumentative matrix assembled by Nietzsche...
  - ...represented a radical **renewal of the 19th century critical project** in Western thought
  - ...became the **mold** for most critical endeavors in the **20th century**

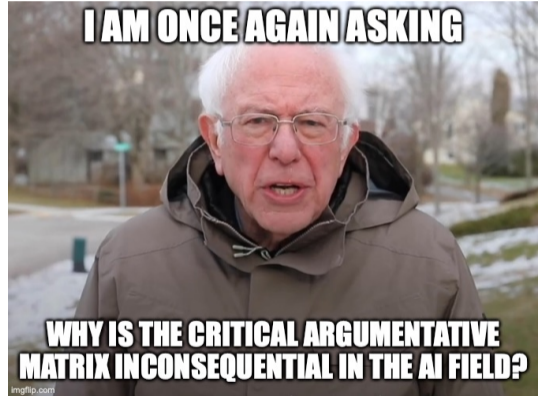
# Criticism and Formalism

- ◇ The argumentative matrix assembled by Nietzsche...
  - ...represented a radical **renewal of the 19th century critical project** in Western thought
  - ...became the **mold** for most critical endeavors in the **20th century**
  - ...seems to contain everything we need to elaborate a critical approach to **AI and LLMs in the 21st century**



# Criticism and Formalism

- ◇ The argumentative matrix assembled by Nietzsche...
  - ...represented a radical **renewal of the 19th century critical project** in Western thought
  - ...became the **mold** for most critical endeavors in the **20th century**
  - ...seems to contain everything we need to elaborate a critical approach to **AI and LLMs in the 21st century**



# The Critical Argumentative Matrix

Knowledge depends on language



The relation between language and the world is essentially arbitrary



Any regularity in language/knowledge is not natural but cultural/social/political



We should resist existing regularities and create new ones

# The Critical Argumentative Matrix

Knowledge depends on language

(Epistemological)



The relation between language and the world is essentially arbitrary



Any regularity in language/knowledge is not natural but cultural/social/political

(Political)



We should resist existing regularities and create new ones

(Aesthetic)

# The Critical Argumentative Matrix

Knowledge depends on language  
(Epistemological)



[The relation between language and the world is essentially arbitrary?]



Any regularity in language/knowledge is not natural but cultural/social/political  
(Political)

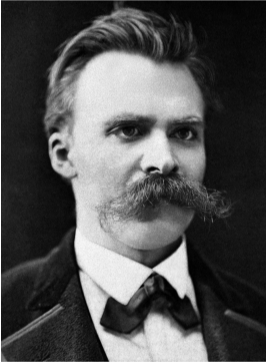


We should resist existing regularities and create new ones  
(Aesthetic)

# Critique and Formalism

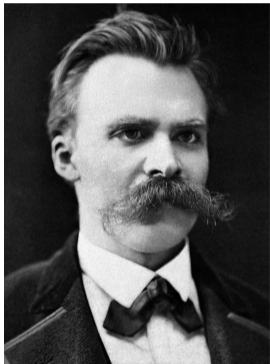
- ◇ At the source of this situation is the new foundational role played by **formal sciences** in the 20th century
  - For a **theory of language**: Carnap, Gödel, Turing, Shannon, Harris, Chomsky...

# Criticism vs Formalism



- ◇ “their eyes merely glide across the surface of things and see ‘forms’; nowhere does their perception lead into truth; instead it is content to receive stimuli and, as it were, to play with its fingers on the back of things.”

# Criticism vs Formalism



- ◇ “their eyes merely glide across the surface of things and see ‘forms’; nowhere does their perception lead into truth; instead it is content to receive stimuli and, as it were, to play with its fingers on the back of things.”
- ◇ “...the great edifice of concepts exhibits the rigid regularity of a Roman columbarium, while logic breathes out that air of severity and coolness which is peculiar to mathematics.”
- ◇ “...if we are forced to comprehend all things under these forms alone, then it is no longer wonderful that what we comprehend in all these things is actually nothing other than these very forms; for all of them must exhibit the laws of number, and number is precisely that which is most astonishing about things.”
- ◇ “...the only things we really know about them [laws of nature] are things which we bring to bear on them: time and space, in other words, relations of succession and number.”

## For a Critical Formalism

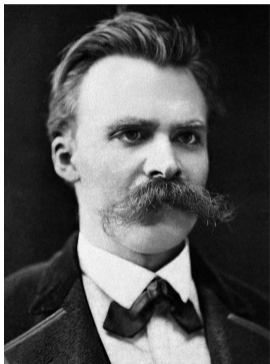
- ◇ At the source of this situation is the new foundational role played by **formal sciences** in the 20th century
  - For a **theory of language**: Carnap, Gödel, Turing, Shannon, Harris, Chomsky...
- ◇ The critical tradition has either **withdrawn** from the areas conquered by formal approaches, or made formal approaches the **target** of criticism



## For a Critical Formalism

- ◇ At the source of this situation is the new foundational role played by **formal sciences** in the 20th century
  - For a **theory of language**: Carnap, Gödel, Turing, Shannon, Harris, Chomsky...
- ◇ The critical tradition has either **withdrawn** from the areas conquered by formal approaches, or made formal approaches the **target** of criticism
- ◇ We need a **new strategy**: Elaborate a **critical formalism**

## For a Critical Formalism



“...all peoples have just such a **mathematically divided firmament of concepts** above them [...] Here, one can certainly **admire humanity** as a mighty **architectural genius** who succeeds in erecting the infinitely complicated cathedral of concepts **on moving foundations**, or even, one might say, **on flowing water**; admittedly, in order to rest on such foundations, it has to be like a thing constructed from cobwebs, **so delicate** that it can be carried off on the waves and yet **so firm** as not to be blown apart by the wind.”

(Nietzsche, 1873)

# For a Critical Formalism

- ◇ In the case of **AI**, a critical formalism can provide:
  - New **epistemological tools** countering dogmatic perspectives stemming from within the field
  - New **theoretical tools** contributing to the non-dogmatic positive production of knowledge

# Outline

Introduction

NLMs as Formal Objects

Conclusion

# Neural LMs as Computable Functions

Neural LM



# Neural LMs as Computable Functions

Neural LM



# Neural LMs as Computable Functions

Neural LM



# Neural LMs as Computable Functions

Neural LM





# Neural LMs as Computable Functions

Neural LM



# Neural LMs as Computable Functions

Neural LM



# Neural LMs as Computable Functions

Neural LM



$f$  !

# Neural LMs as Computable Functions

Neural LM



Function

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

# Neural LMs as Computable Functions

Neural LM

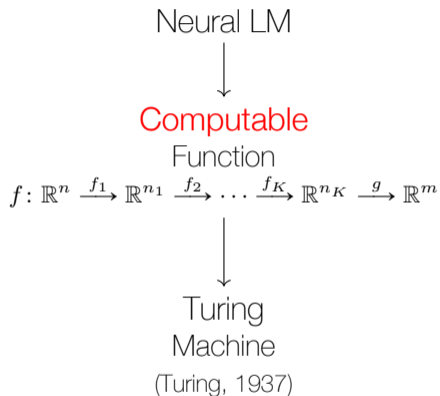


Computable

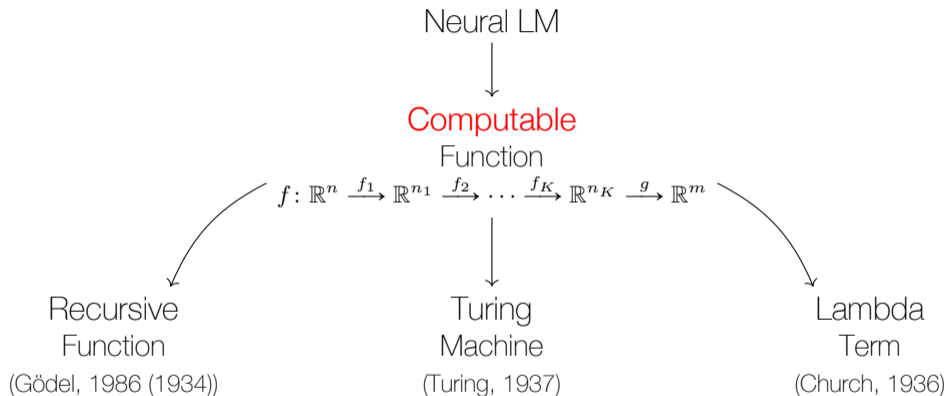
Function

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

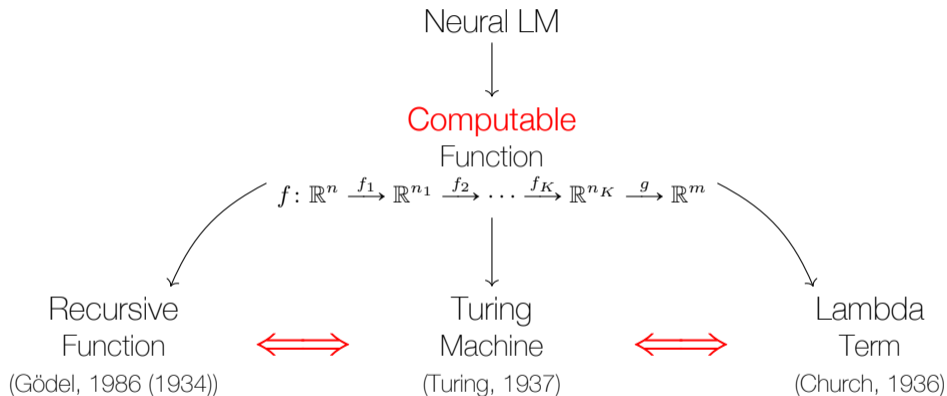
# Neural LMs as Computable Functions



# Neural LMs as Computable Functions

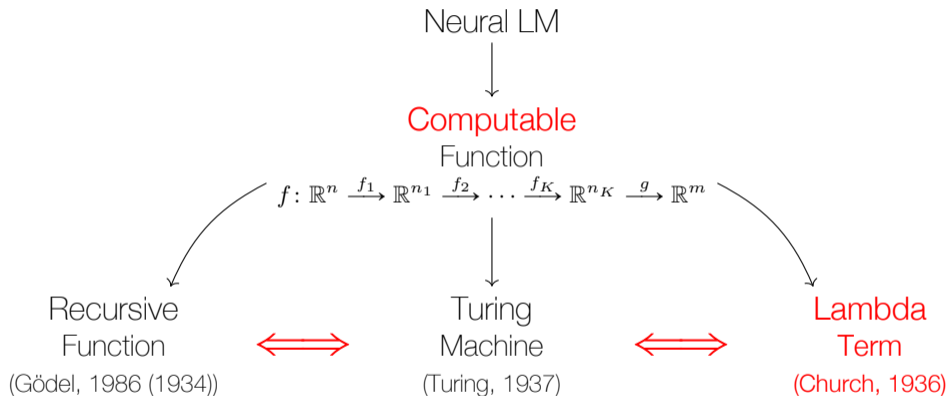


# Neural LMs as Computable Functions

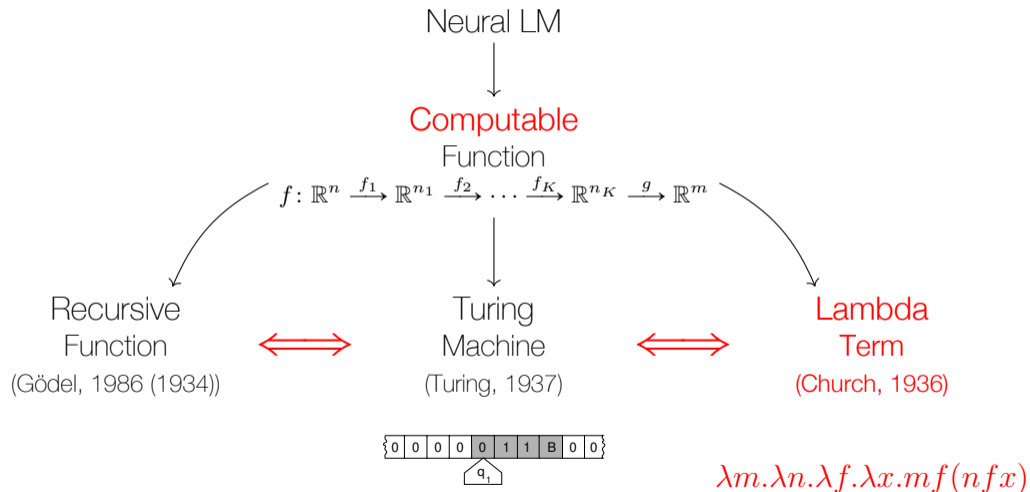




# Neural LMs as Computable Functions



# Neural LMs as Computable Functions



credit: Nynexman4464

## $\beta$ -reduction in $\lambda$ -calculus

$yxz$

## $\beta$ -reduction in $\lambda$ -calculus

$\lambda x.yxz$

## $\beta$ -reduction in $\lambda$ -calculus

$$(\lambda x. yxz)t$$

## $\beta$ -reduction in $\lambda$ -calculus

$(\lambda x. yxz)t$

$\Downarrow$

$ytz$

## Empirical Evaluation

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

## Empirical Evaluation

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

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

...

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



# Empirical Evaluation

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

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

...

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

# Empirical Evaluation

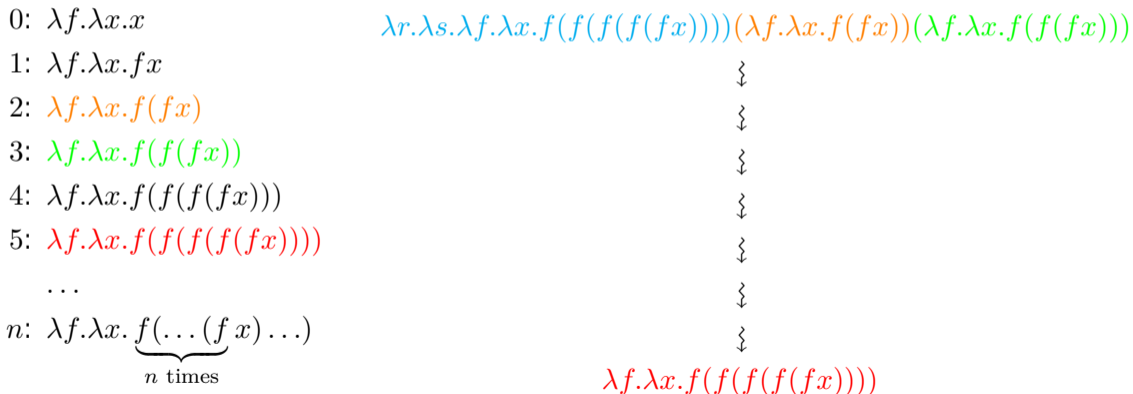
$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$

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

# Empirical Evaluation

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

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



# Interpretability

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

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

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

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

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

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

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

...

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

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

⋮

⋮

⋮

⋮

⋮

⋮

⋮

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

# Interpretability

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

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

...

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

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

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

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

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

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

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

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

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

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

# Interpretability

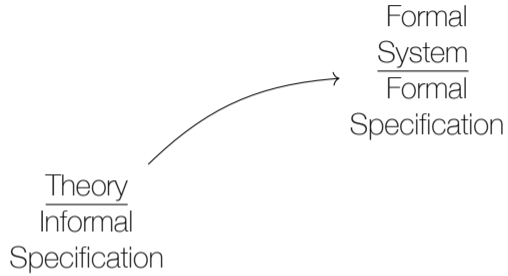
$P := \lambda m. \lambda n. \lambda f. \lambda x. mf(nfx)$

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

# Making it Explicit

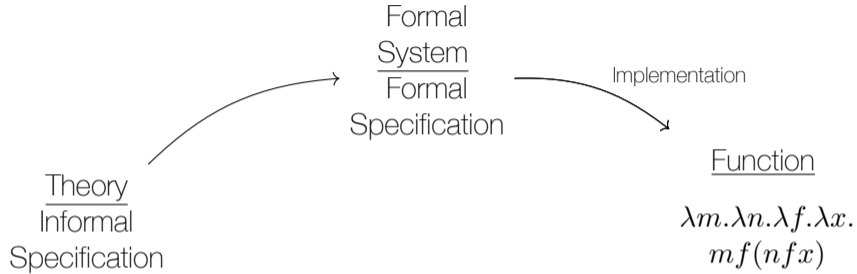
Theory  
Informal  
Specification

# Making it Explicit

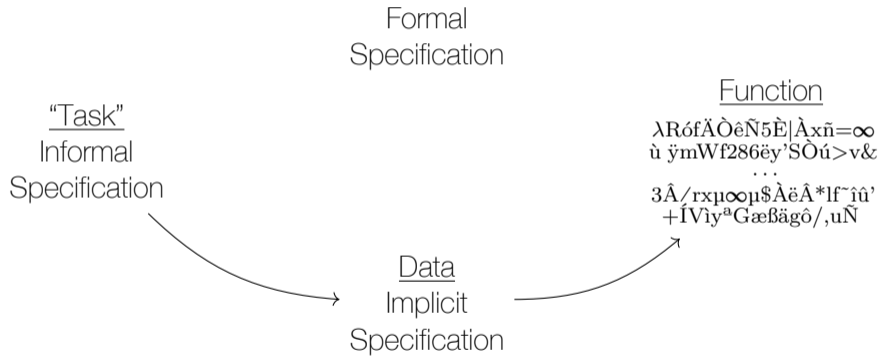




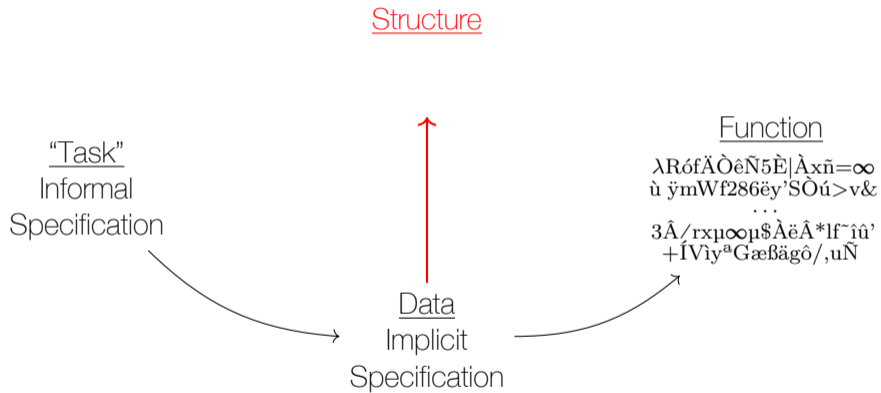
# Making it Explicit



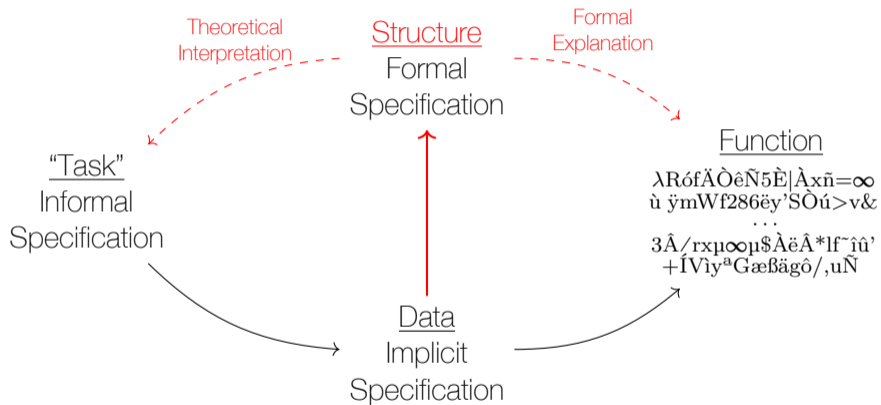
# Making it Explicit



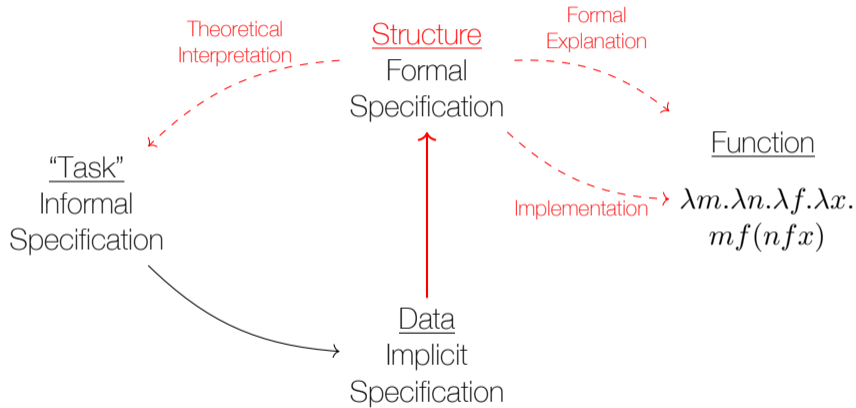
# Making it Explicit



# Making it Explicit



# Making it Explicit



# Outline

Introduction

NLMs as Formal Objects

Conclusion

## Conclusion: For a Critical Formalism

- ◇ It is urgent to address the **epistemological** dimension of the critical project in its own terms
- ◇ This requires to develop a **critical approach within formal sciences** where formalization is not assumed to lead to **naturalization**
  - The new role of **data** within formal sciences is crucial in this sense
- ◇ A **critical formalism** will be incomplete if it remains disconnected from the **political**, and even the **artistic** dimension of the critical program
  - We need a **new alliance** between the **formal sciences**, the **human sciences**, and the **arts**.

# Reference Papers

- ◇ Gastaldi, J. L. (2021). Why Can Computers Understand Natural Language? *Philosophy & Technology*, 34(1), 149–214. <https://doi.org/10.1007/s13347-020-00393-9>
- ◇ Gastaldi, J. L., & Pellissier, L. (2021). The calculus of language: Explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*. <https://doi.org/10.1080/03080188.2021.1890484>
- ◇ Bradley, T.-D., Gastaldi, J. L., & Terilla, J. (2024). The structure of meaning in language: Parallel narratives in linear algebra and category theory. *Notices of the American Mathematical Society*. <https://api.semanticscholar.org/CorpusID:263613625>



# References I

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Bradley, T.-D., Gastaldi, J. L., & Terilla, J. (2024). The structure of meaning in language: Parallel narratives in linear algebra and category theory. *Notices of the American Mathematical Society*. <https://api.semanticscholar.org/CorpusID:263613625>
- Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2), 345–363.
- Gastaldi, J. L. (2021). Why Can Computers Understand Natural Language? *Philosophy & Technology*, 34(1), 149–214. <https://doi.org/10.1007/s13347-020-00393-9>
- Gastaldi, J. L., & Pellissier, L. (2021). The calculus of language: Explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*. <https://doi.org/10.1080/03080188.2021.1890484>
- Gödel, K. (1986 (1934)). On undecidable propositions of formal mathematical systems. In *Collected works* (pp. 346–371). Clarendon Press Oxford University Press.
- Kirschenbaum, M. (2023). *Again theory: A forum on language, meaning, and intent in the time of stochastic parrot*. <https://critinq.wordpress.com/2023/06/26/again-theory-a-forum-on-language-meaning-and-intent-in-the-time-of-stochastic-parrots/>

## References II

- Nietzsche, F. (1873). On truth and lying in a non-moral sense [Originally unpublished; written in 1873.]. (R. Speirs, Trans.). In R. Geuss & R. Speirs (Eds.), *The birth of tragedy and other writings* (pp. 141–153). Cambridge University Press.
- 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>
- Underwood, T. (2023, October 15). *The empirical triumph of theory* [Accessed: 2023-10-15]. <https://critinq.wordpress.com/2023/06/29/the-empirical-triumph-of-theory/>

ITS, CUNY Graduate Center  
New York, NY, USA

*Toward a Critical Formalism*  
From the Philosophy to the Theory  
of Distributional Language Models

Juan Luis Gastaldi

**ETH** zürich

April 11, 2025