

# NaturalProver

**Sean Welleck**

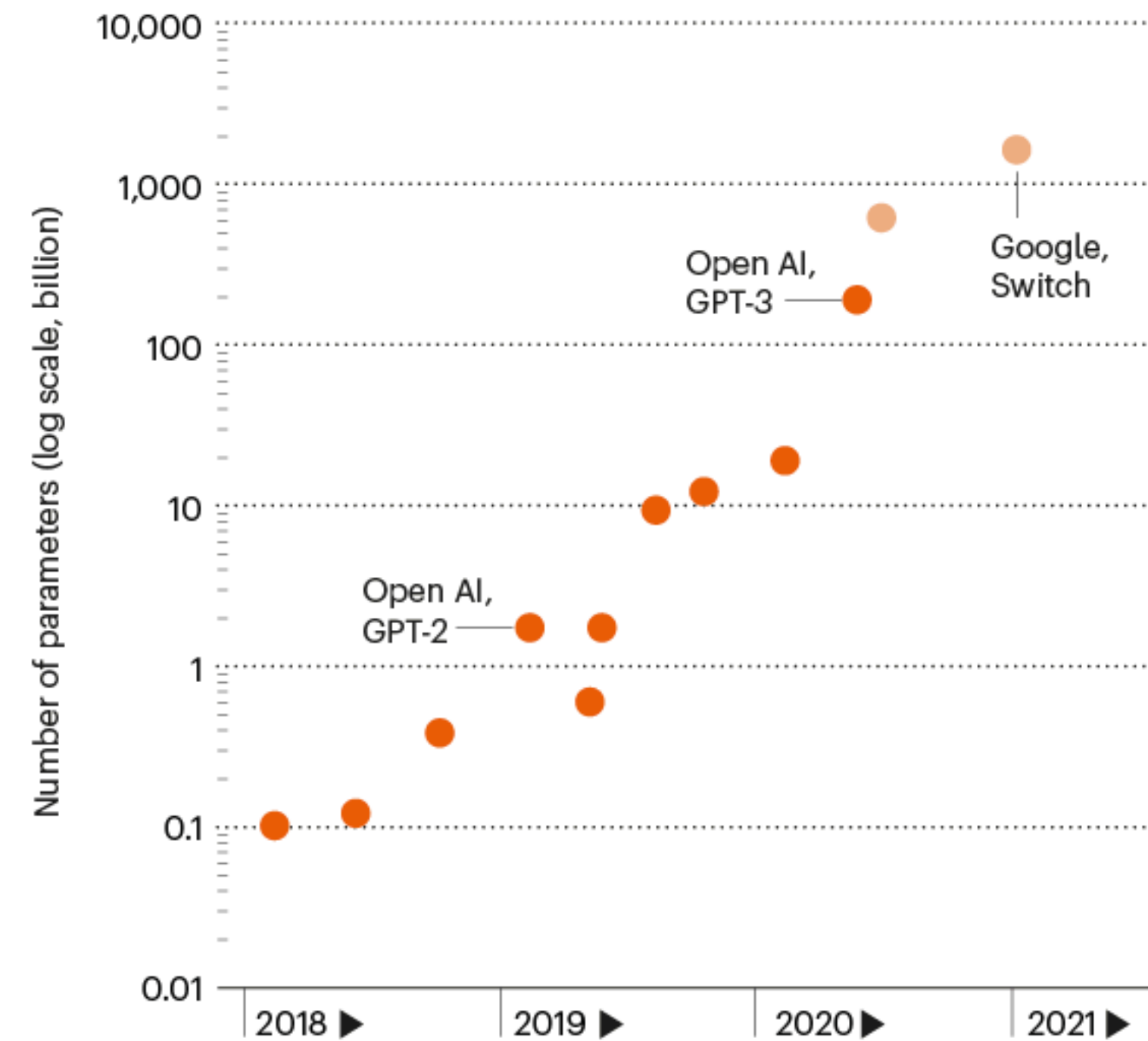
Joint work with Jiacheng Liu, Ximing Lu, Hanna Hajishirzi, Yejin Choi

# Neural Language Models

## LARGER LANGUAGE MODELS

The scale of text-generating neural networks is growing exponentially, as measured by the models' parameters (roughly, the number of connections between neurons).

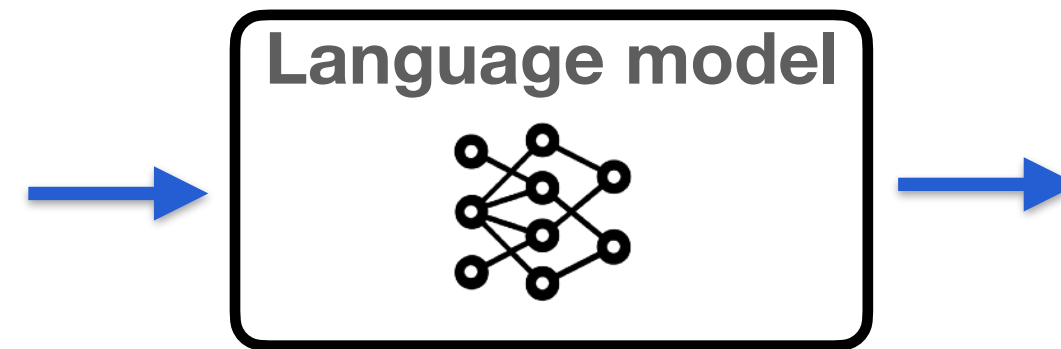
● 'Dense' models ● 'Sparse' models\*



\*Google's 1.6-trillion parameter 'sparse' model has performance equivalent to that of 10 billion to 100 billion parameter 'dense' models. ©nature

# Neural Language Models

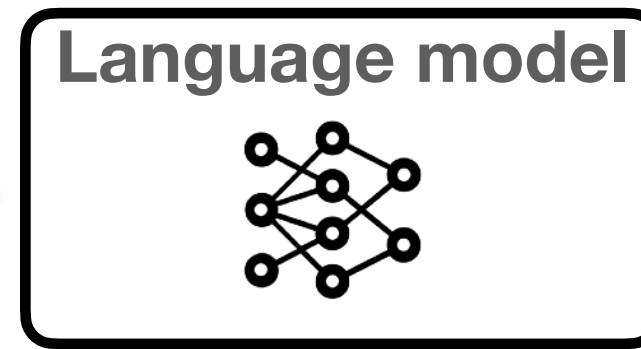
What is the meaning  
of life?



The meaning of life is a question  
that has been asked by people  
throughout history.  
There is no one correct answer to  
this question.

# Neural Language Models

```
theorem difficult_theorem : a > 0 ∧ b > 0 → a*b > 0 :=  
begin  
  rintros ⟨h1, h2⟩,  
  gptf,  
end
```



```
Try this: exact mul_pos h1 h2  
All predictions:  
-----  
Try this: exact lt_of_le_of_ne (mul_le_mul (le_refl a) (le_of_lt h1)) (mul_le_mul (le_of_lt h1) (le_of_lt h2))  
Try this: exact lt_of_le_of_pos (mul_le_mul (le_of_lt h1) (le_of_lt h2)) (mul_pos h1 h2)  
Try this: exact mul_pos h1 h2  
Try this: exact lt_of_le_of_ne (mul_le_mul_of_nonneg_right (mul_le_mul (le_of_lt h1) (le_of_lt h2)) (le_of_lt h2)) (mul_pos h1 h2)  
Try this: refine lt_of_le_of_lt (mul_le_mul (le_of_eq (pos_of_neg (neg_pos_of_neg (pos_neg.2 h2))) h2)) (mul_pos h1 h2)  
Try this: PGame.mk_ => nat.mul_pos_neg_of_neg_of_neg  
Try this: exact lt_of_pos_of_neg (mul_pos h1 h2)  
Try this: exact mul_pos (neg_pos_of_neg (pos_neg.2 h2)) h2
```



# Informal & formal mathematics

Informal ←

### Textbook

## 6.2. The Intermediate-Value Theorem

The intermediate-value theorem states that if  $f$  is a continuous function on the closed interval  $[a, b]$  and if  $d$  is between  $f(a)$  and  $f(b)$ , then there is a number  $c$  between  $a$  and  $b$  such that  $f(c) = d$ . The power of the theorem lies in the fact that it provides a way of knowing about the existence of something without requiring that it be explicitly found.

As an example, let us show that  $-2x^5 + 4x = 1$  has a solution in the interval  $(0, 1)$ . Consider  $f(x) = -2x^5 + 4x$ , and take two “pot-shots”:  $f(0)$  is too small, and  $f(1)$  is too large. Therefore, by the intermediate-value theorem, there is a number in  $(0, 1)$  that is just right.

### Proofwiki

#### Theorem

Let  $f : S \rightarrow \mathbb{R}$  be a [real function](#) on some [subset](#)  $S$  of  $\mathbb{R}$ .  
 Let  $I \subseteq S$  be a [real interval](#).  
 Let  $f : I \rightarrow \mathbb{R}$  be [continuous](#) on  $I$ .  
 Then  $f$  is a [Darboux function](#).

That is:  
 Let  $a, b \in I$ .

Let  $k \in \mathbb{R}$  lie between  $f(a)$  and  $f(b)$ .  
 That is, either:  
 $f(a) < k < f(b)$   
 or:  
 $f(b) < k < f(a)$

Then  $\exists c \in (a, b)$  such that  $f(c) = k$ .

Formal →

### Mathlib/Lean

```
/-- **Intermediate Value Theorem** for continuous functions on c
`f a ≤ t ≤ f b`.-/
lemma intermediate_value_Icc {a b : α} (hab : a ≤ b) {f : α → δ}
  Icc (f a) (f b) ≤ f '' (Icc a b) :=
  is_preconnected_Icc.intermediate_value (left_mem_Icc.2 hab) (right)
```

### Isabelle/HOL

```
subsection <Intermediate Value Theorem>
lemma IVT':
  fixes f :: "'a::linear_continuum_topology ⇒ 'b::linorder_topology"
  assumes y: "f a ≤ y" "y ≤ f b" "a ≤ b"
  and *: "continuous_on {a .. b} f"
  shows "∃x. a ≤ x ∧ x ≤ b ∧ f x = y"
proof -
  have "connected {a..b}"
  unfolding connected_iff_interval by auto
  from connected_continuous_image[OF * this, THEN connectedD_interval, of "f a" "f b" y] y
  show ?thesis
  by (auto simp add: atLeastAtMost_def atLeast_def atMost_def)
qed
```

# Interesting for machine learning

## Problem Solving

Equation	Solution
$x^2 = \frac{16x^2 - 12x^2 + 2x}{(-16x^2 + 112x^2 - 204x^2 + 28x^2 - x^4 + 1)^{1/2}}$	
$3xy \cos(x) - \sqrt{9x^2 \sin(x)^2 + 1}y' + 3y \sin(x) = 0$	
$4x^2y^2 - 8x^2y^3 - 8x^2y^4 - 3x^2y^5 - 8x^2y^6 - 8x^2y^7 - 3x^2y^8 - 5xy^9 - 3y^{10} = 0$	

### Grade School Math

[Blog Post] [Paper]

State-of-the-art language models can match human performance on many tasks, but they still perform multi-step mathematical reasoning. To diagnose the failures of current models and see how we're releasing GSM8K, a dataset of 8.5K high quality linguistically diverse grade school math problems, we find that even the largest transformer models fail to achieve high test performance, despite the simplicity of this problem distribution.

[Lample & Charton 2019]

[Cobbe et al 2021]

## Informal Theorem Proving

### Theorem Co-Countable Measure is Probability Measure

Let  $X$  be an uncountable set.

Let  $\mathcal{A}$  be the  $\sigma$ -algebra of countable sets on  $X$ .

Then the co-countable measure  $\mu$  on  $X$  is a probability measure.

### Proof

By Co-Countable Measure is Measure,  $\mu$  is a measure.

By Relative Complement with Self is Empty Set, have  $X^C(X) = \emptyset$ .

As  $\emptyset$  is countable, it follows that  $X$  is co-countable.

Hence  $\mu(X) = 1$ , and so  $\mu$  is a probability measure.

## Long-form QA

**Question** Why did we decide that certain words were "bad" and shouldn't be used in social settings?

**Answer** We have all been taught that there are certain words that should not be used in public, at least not without good reason. But why do some words fall into this category while others do not? It turns out that there is not a single answer to this question, as each word has a different history and carries different meanings in different cultures and families [2]. However, there are some general categories that "bad" words tend to fall into. "Bad" words generally relate to parts of life that we don't like talking about in public, like bathroom functions, or negative ways of talking about people's religion, skin color, ability, or gender [3]. Often, people say bad words when they are experiencing strong emotions, and those words are linked to our emotions [3].

**References (titles only)** [1, 2, 3] Why Are Some Words 'Bad'? | Vermont Public Radio (www.vpr.org)  
[4] On Words: 'Bad' Words and Why We Should Study Them | UVA Today (news.virginia.edu)  
[5] The Science of Curse Words: Why The &@#! Do We Swear? (www.babbel.com)

[Nakano et al 2022]

## Formal Theorem Proving

```
theorem difficult_theorem :
begin
  rintros (h1, h2),
  gptf,
end
```

Solving (Some) Formal Math Olympiad Problems

We built a neural theorem prover for Lean that learned to solve a variety of challenging high-school olympiad problems, including problems from the AMC12 and AIME competitions, as well as two problems adapted from the IMO. The prover uses a language model to find proofs of formal statements. Each time we find a new proof, we use it as new training data, which improves the neural network and enables it to iteratively find solutions to harder and harder statements.

Try this: PGame.mk => nat.mul\_pos\_neg\_of\_neg\_of\_neg

Try this: exact lt\_of\_pos\_of\_neg (mul\_pos h1 h2)

Try this: exact mul\_pos (neg\_pos\_of\_neg (pos\_neg.2 h2)) h1

[Han et al 2021] [Polu et al 2022]

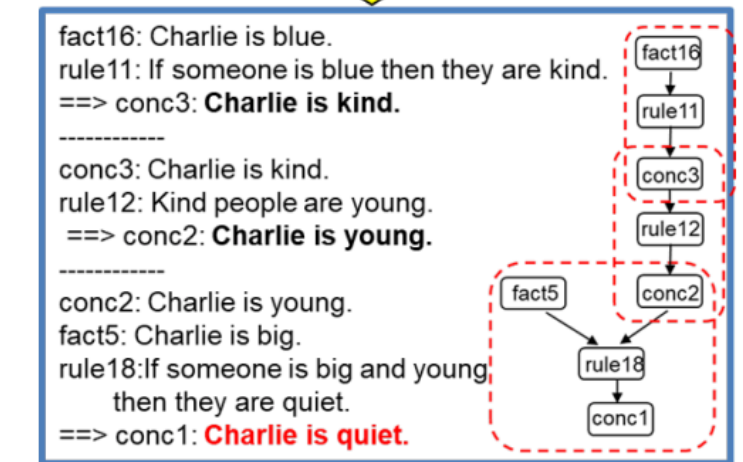
## Systematic Reasoning with Language

### Theory (Facts + Rules)

fact1: Erin is young.  
...  
fact5: Charlie is big.  
fact6: Dave is white.  
...  
rule10: If someone is young and not kind then they are big.  
rule11: If someone is blue then they are kind.  
rule12: Kind people are young.  
...  
fact16: Charlie is blue.  
...  
rule18: If someone is big and young then they are quiet.

q: Charlie is quiet?

ProofWriter



Generated Proof

[Tajford et al 2020]

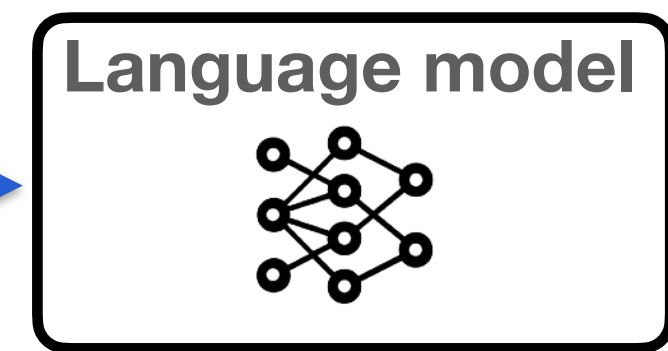


# Interesting applications

► Formal mathematics: gpt-f tactic [Han et al 2021]

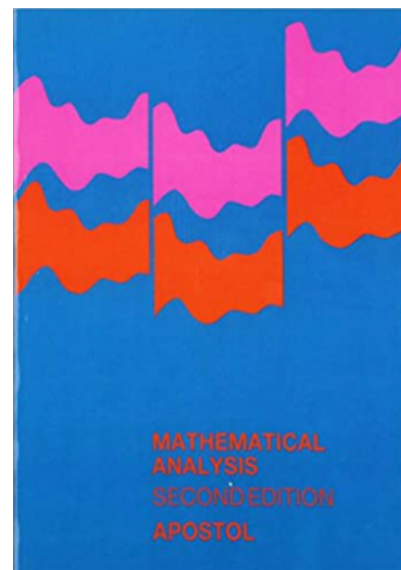
► Code: GitHub CoPilot / Codex

```
theorem difficult_theorem : a > 0 ∧ b > 0 → a*b > 0 :=
begin
  rintros ⟨h1, h2⟩,
  gptf,
end
```

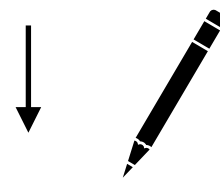


```
Try this: exact mul_pos h1 h2
All predictions:
-----
Try this: exact lt_of_le_of_ne (mul_le_mul (le_refl a) (le_of_lt h1)) (mul_pos h1 h2)
Try this: exact lt_of_le_of_pos (mul_le_mul (le_of_lt h1) (le_of_pos h2)) (mul_pos h1 h2)
Try this: exact mul_pos h1 h2
Try this: exact lt_of_le_of_ne (mul_le_mul_of_nonneg_right (le_of_pos h1) (le_of_pos h2)) (mul_pos h1 h2)
Try this: refine lt_of_le_of_lt (mul_le_mul (le_of_eq (pos_of_neg h1)) (le_of_pos h2)) (mul_pos h1 h2)
Try this: PGame.mk_ => nat.mul_pos_neg_of_neg_of_neg (mul_pos h1 h2)
Try this: exact lt_of_pos_of_neg (mul_pos h1 h2)
Try this: exact mul_pos (neg_pos_of_neg (pos_neg.2 h2)) h2
```

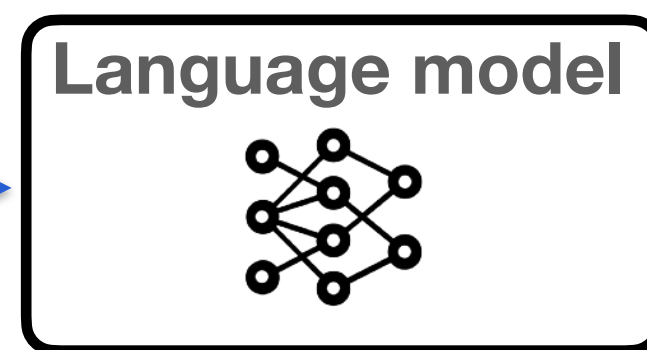
► Mathematics



**5.14** Given a function  $f$  defined and having a finite derivative  $f'$  in the half-open interval  $0 < x \leq 1$  and such that  $|f'(x)| < 1$ . Define  $a_n = f(1/n)$  for  $n = 1, 2, 3, \dots$ , and show that  $\lim_{n \rightarrow \infty} a_n$  exists. *Hint.* Cauchy condition.



**Proof:** Consider  $n \geq m$ , and by Mean Value Theorem,



$$|a_n - a_m| = |f(1/n) - f(1/m)|$$

$$|a_n - a_m| = |f(1/n) - f(1/m)| = |f'(p)| \left| \frac{1}{n} - \frac{1}{m} \right| \leq \left| \frac{1}{n} - \frac{1}{m} \right|$$

then  $\{a_n\}$  is a Cauchy sequence since  $\{1/n\}$  is a Cauchy sequence. Hence, we know that  $\lim_{n \rightarrow \infty} a_n$  exists.



# Outline

- Language models for generating proofs and suggesting next steps
  - Premise-augmented generation
  - Premise constraints
- Evaluation schema & study
- Interactive human-machine proving

# Data: NaturalProofs-Gen

- ▶ **NaturalProofs** [Welleck et al 2021]: Multi-domain dataset of mathematical proofs
- ▶ **ProofWiki domain**: 20k theorems, 12.5k definitions

**NaturalProofs**

- Broad-coverage
- Deep-coverage
- Low-resource

**ProofWiki**

quod erat demonstrandum

Main Page  
Community discussion  
Community portal  
Recent changes  
Random proof  
Help  
FAQ  
ProofWiki  $L^{\infty}$  commands

ProofWiki.org

Proof Index  
Definition Index  
Symbol Index  
Axiom Index  
Mathematicians  
Books  
Sandbox  
All Categories  
Glossary  
Jokes

To Do

Proofread Articles  
Wanted Proofs  
More Wanted Proofs  
Help Needed  
Research Required  
Stub Articles  
Tidy Articles  
Improvements Invited  
Refactoring  
Missing Links  
Maintenance

Tools

What links here  
Related changes  
Special pages  
Printable version  
Permanent link  
Page information

## NaturalProofs | Proofwiki

Main page [Discussion](#) [Read](#) [View sou](#)

### Welcome to ProofWiki

**ProofWiki** is an online compendium of mathematical proofs! Our goal is the collection, collaboration and classification of mathematical proofs. If you feel free to [register for an account](#). Thanks and enjoy!

If you have any questions, comments, or suggestions please post on the [discussion](#) page, or contact one of the [administrators](#). Also, feel free to be the first with your idea.

To see what's currently happening in the community, visit the [community portal](#).

23,703 Proofs — 18,544 Definitions — Help

[Follow @ProofWiki](#)

### Featured Proof

#### Laplace Transform of Complex Power

#### Theorem

Let  $q$  be a constant complex number with  $\Re(q) > -1$

Let  $f^q : \mathbb{R}_{>0} \rightarrow \mathbb{C}$  be a branch of the complex power multifunction chosen such that  $f$  is continuous on the half-plane  $\Re(s) > 0$ .

Then  $f$  has a Laplace transform given by:

$$\mathcal{L}\{f^q\} = \frac{\Gamma(q+1)}{s^{q+1}}$$

where  $\Gamma$  denotes the gamma function.

#### Proof

By definition of Laplace transform for a function not continuous at zero

$$\mathcal{L}\{f^q\} = \lim_{\epsilon \rightarrow 0^+} \lim_{L \rightarrow +\infty} I(\epsilon, L)$$

where:

$$r^L$$

# Data: NaturalProofs-Gen

Theorem **x**

## Co-Countable Measure is Probability Measure

---

### Theorem

---

Let  $X$  be an **uncountable set**.

Let  $\mathcal{A}$  be the  **$\sigma$ -algebra of countable sets** on  $X$ .

Then the **co-countable measure**  $\mu$  on  $X$  is a **probability measure**.

Proof **y**

### Proof

---

By **Co-Countable Measure is Measure**,  $\mu$  is a **measure**.

By **Relative Complement with Self is Empty Set**, have  $\mathbb{C}_X(X) = \emptyset$ .

As  $\emptyset$  is **countable**, it follows that  $X$  is **co-countable**.

Hence  $\mu(X) = 1$ , and so  $\mu$  is a **probability measure**.

■

# Data: NaturalProofs-Gen

## Theorem $\mathbf{x}$

### Co-Countable Measure is Probability Measure

#### Theorem

Let  $X$  be an **uncountable set**.

Let  $\mathcal{A}$  be the  **$\sigma$ -algebra of countable sets** on  $X$ .

Then the **co-countable measure**  $\mu$  on  $X$  is a **probability measure**.

Proof  $\mathbf{y} = (y_1, \dots, y_{T_y})$

- $y_t$  : 'proof step'

#### Proof

By **Co-Countable Measure is Measure**,  $\mu$  is a **measure**.

By **Relative Complement with Self is Empty Set**, have  $\mathbb{C}_X(X) = \emptyset$ .

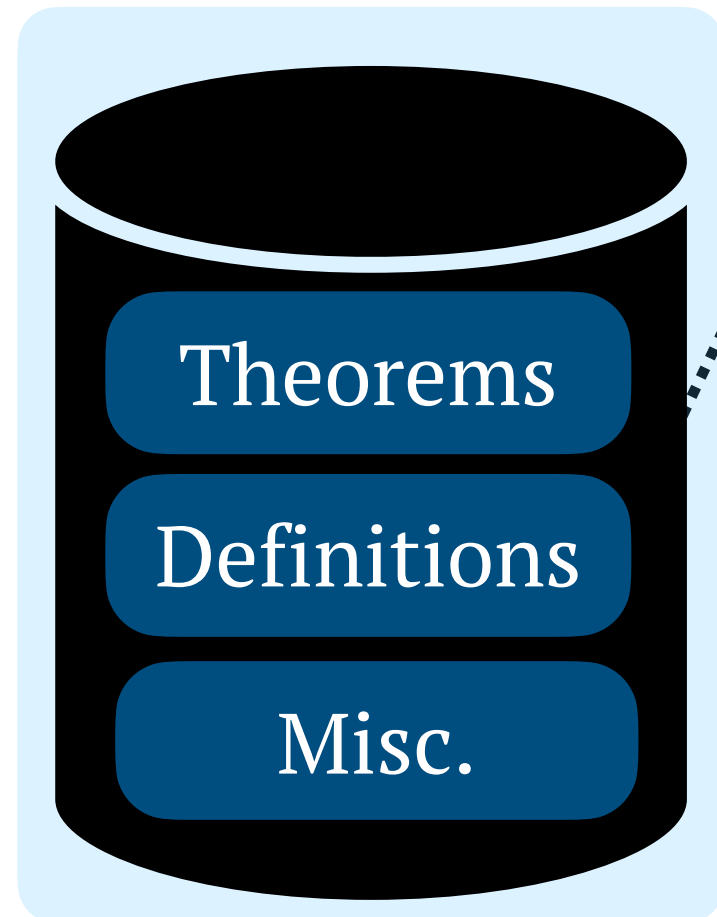
As  $\emptyset$  is **countable**, it follows that  $X$  is **co-countable**.

Hence  $\mu(X) = 1$ , and so  $\mu$  is a **probability measure**.

■



# Data: NaturalProofs-Gen



~30k references  
(theorems, definitions)

## Relative Complement with Self is Empty Set

The **relative complement** of a **set** in itself is the **empty set**:

$$\complement_S(S) = \emptyset$$

## Definition:Co-Countable Measure

Let  $X$  be an **uncountable set**.

Let  $\Sigma$  be the  **$\sigma$ -algebra of countable sets** on  $X$ .

Then the **co-countable measure (on  $X$ )** is the **measure**  $\mu : \Sigma \rightarrow \overline{\mathbb{R}}$  defined as:

$$\forall E \in \Sigma : \mu(E) := \begin{cases} 0 & \text{if } E \text{ is countable} \\ 1 & \text{if } E \text{ is co-countable} \end{cases}$$

where:

$\overline{\mathbb{R}}$  denotes the **extended real numbers**

$E$  is **co-countable** if and only if  $X \setminus E$  is countable.

## Proof

By **Co-Countable Measure is Measure**,  $\mu$  is a **measure**.

By **Relative Complement with Self is Empty Set**, have  $\complement_X(X) = \emptyset$ .

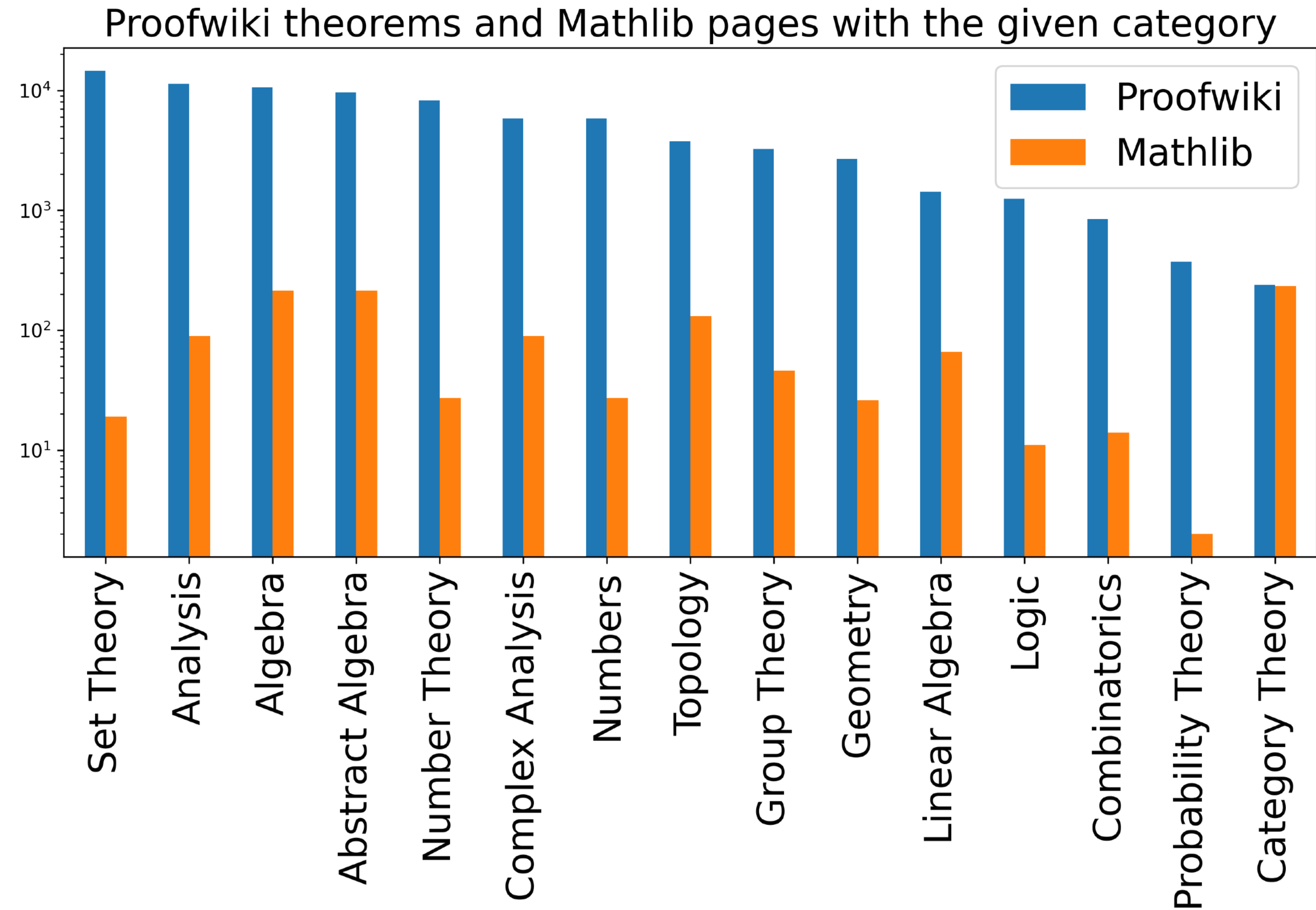
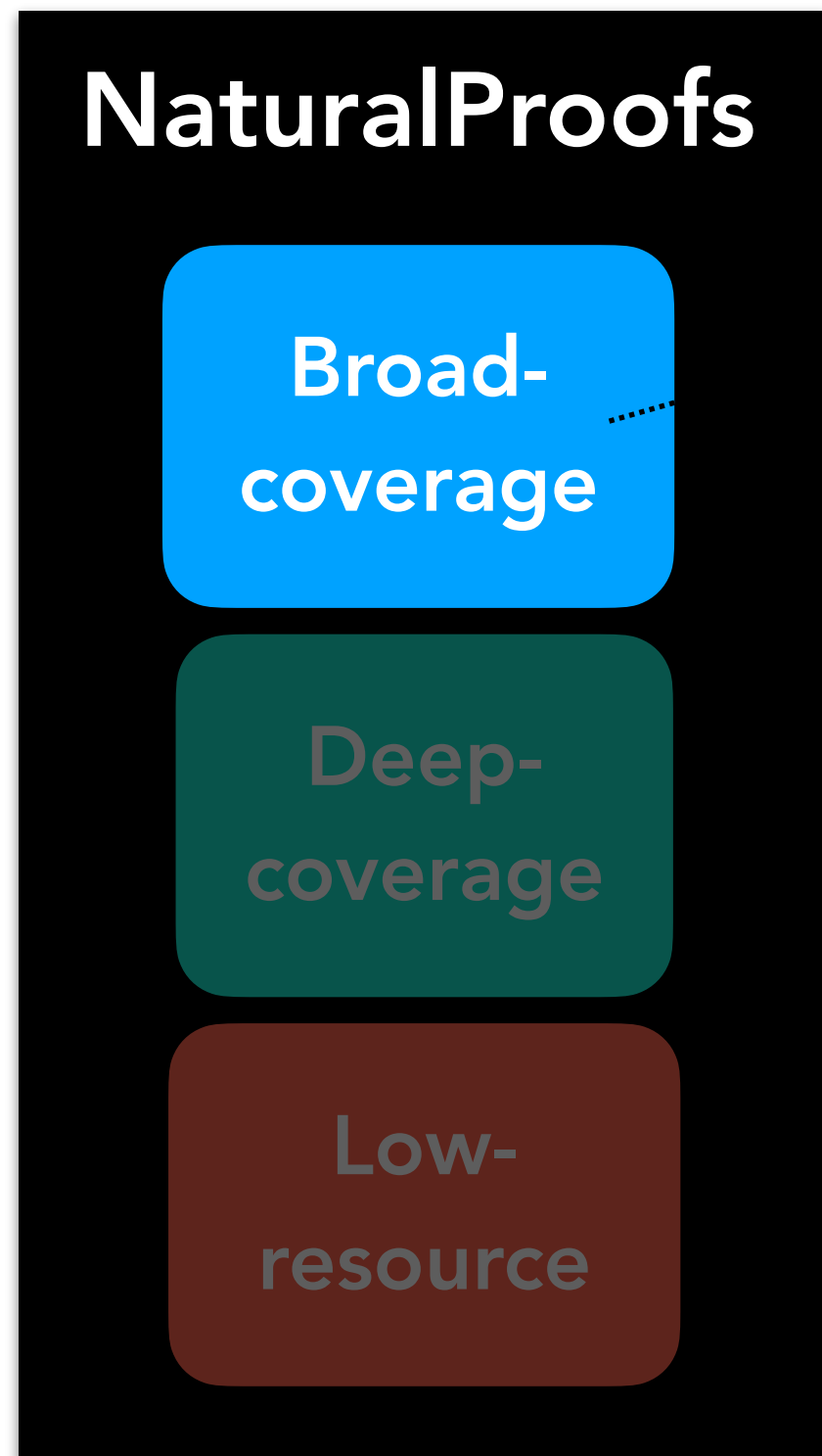
As  $\emptyset$  is **countable**, it follows that  $X$  is **co-countable**.

Hence  $\mu(X) = 1$ , and so  $\mu$  is a **probability measure**.

■

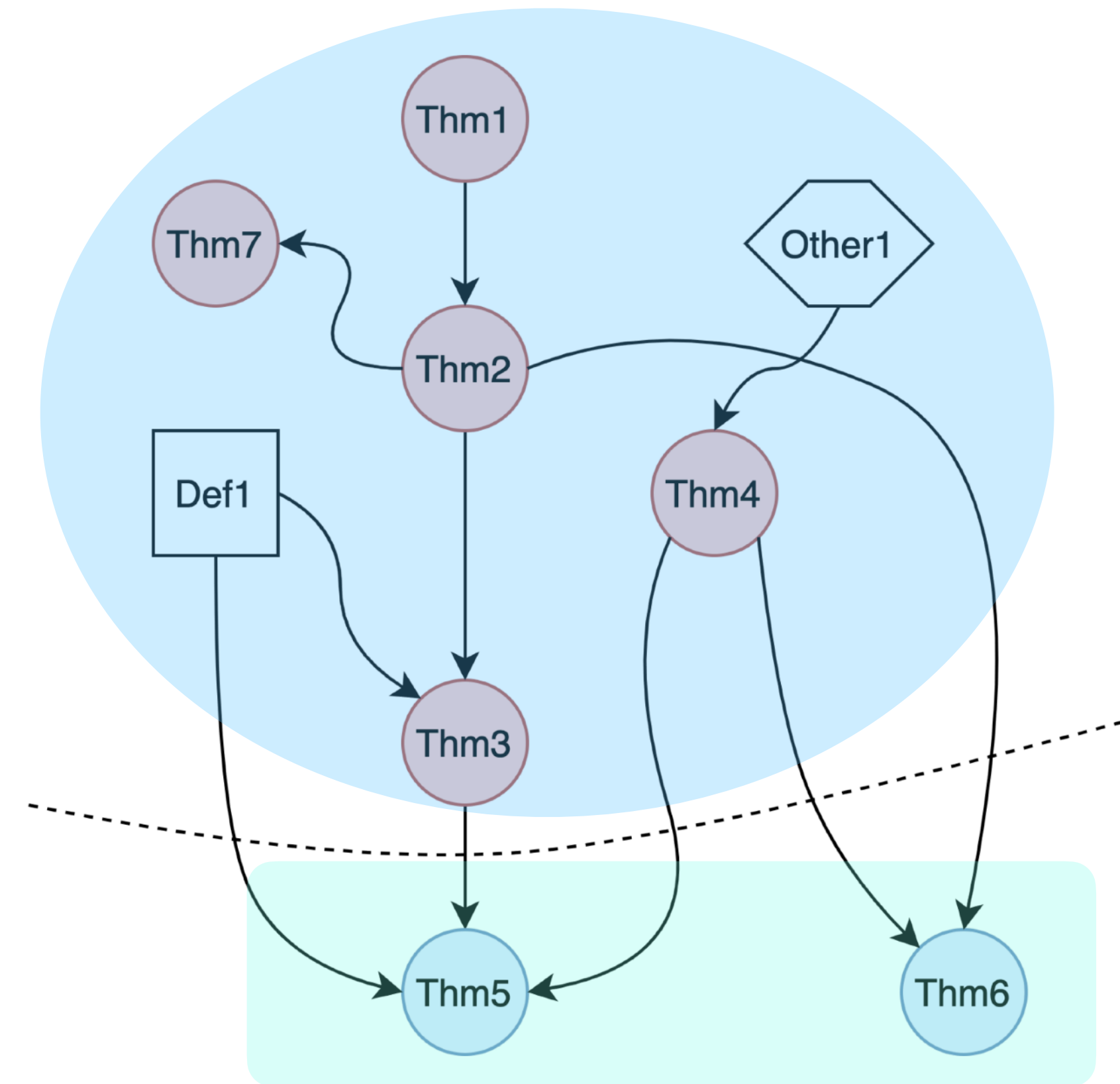


# Data: NaturalProofs-Gen



# Data: NaturalProofs-Gen

- ▶ **Train:** 12.5k theorems/proofs
- ▶ **Validation/Test:** 1k theorems/proofs each
  - ▶ **Evaluation:** 100 valid, 100 test theorems



# Proof generation task

## Input

Even Integers form Commutative Ring

---

Let  $2\mathbb{Z}$  be the [set of even integers](#).

Then  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#).

However,  $(2\mathbb{Z}, +, \times)$  is not an [integral domain](#).

Theorem **x**

## Output

From [Integer Multiples form Commutative Ring](#),  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#).

As  $2 \neq 1$ , we also have from [Integer Multiples form Commutative Ring](#) that  $(2\mathbb{Z}, +, \times)$  has no [unity](#).

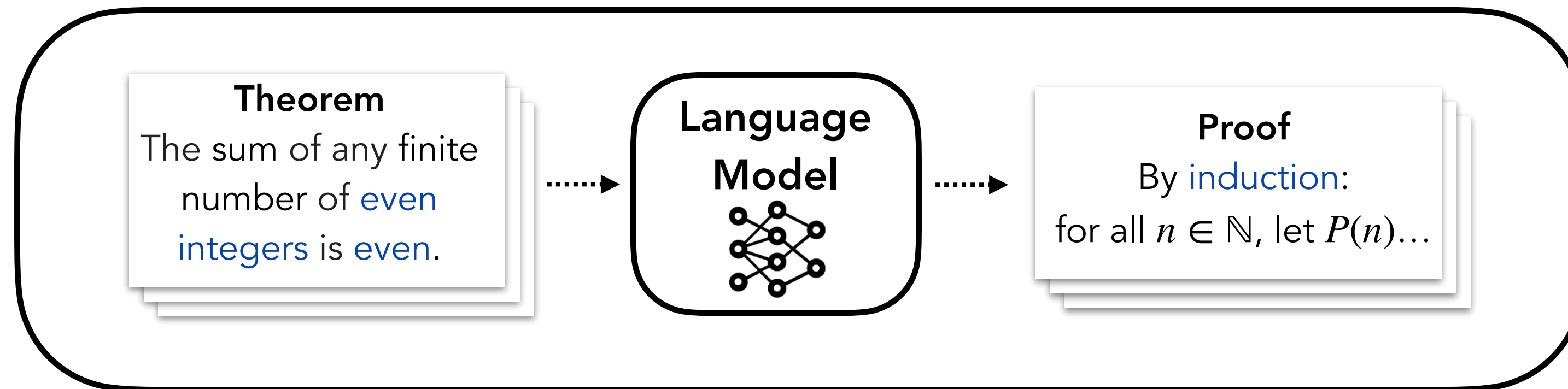
Hence by definition it is not an [integral domain](#).

Full proof **y**

# Vanilla language model

**Training** Finetune on (theorem, proof) pairs

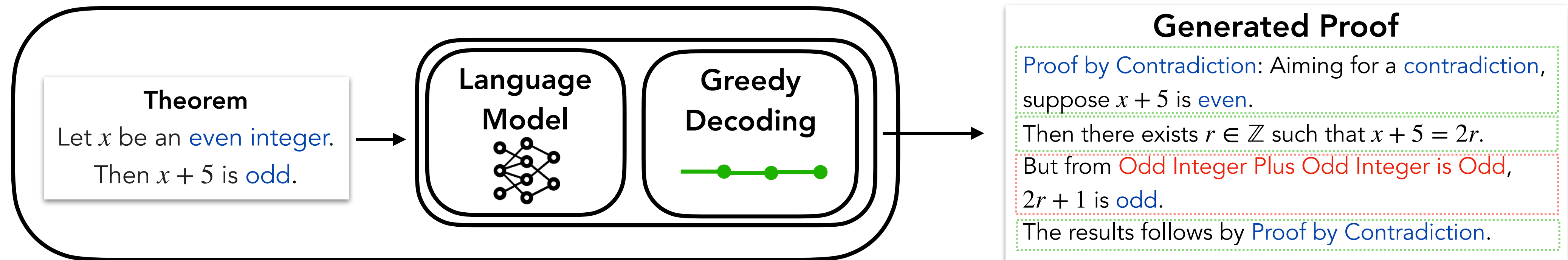
$$\max_{\theta} \sum_{\mathbf{x}, \mathbf{y} \in D_{train}} \log p_{\theta}(\mathbf{y} | \mathbf{x})$$



# Vanilla language model

**Generation** Greedily-decode a proof

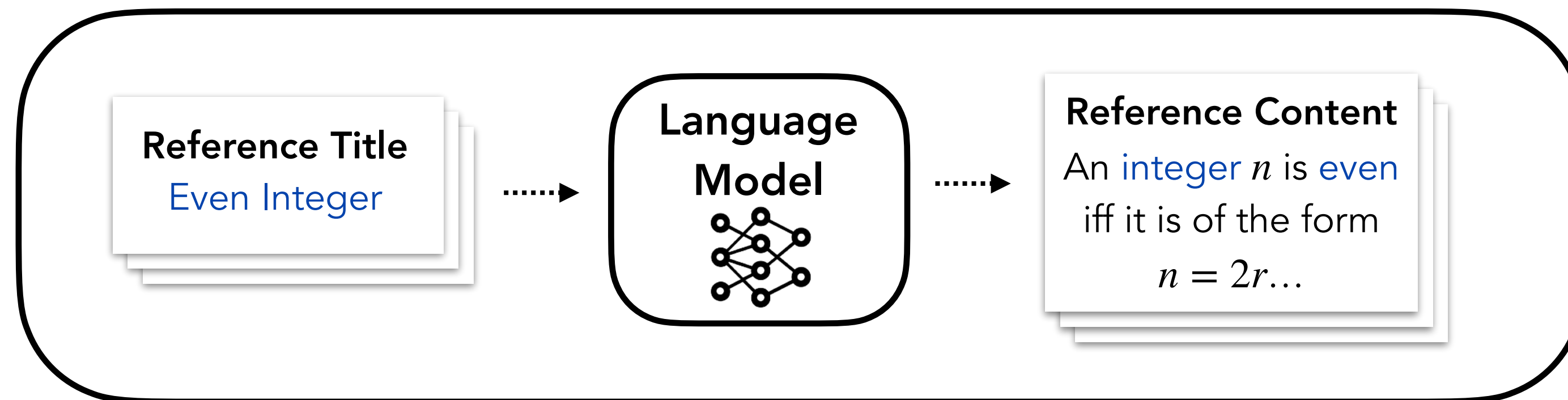
$$\hat{\mathbf{y}} = \text{greedy}(p_{\theta}(\mathbf{y} | \mathbf{x}))$$



# NaturalProver

Condition on background knowledge (*references R*)  $p_{\theta}(\mathbf{y} \mid \mathbf{x}, R)$

## 1. Reference-reconstruction

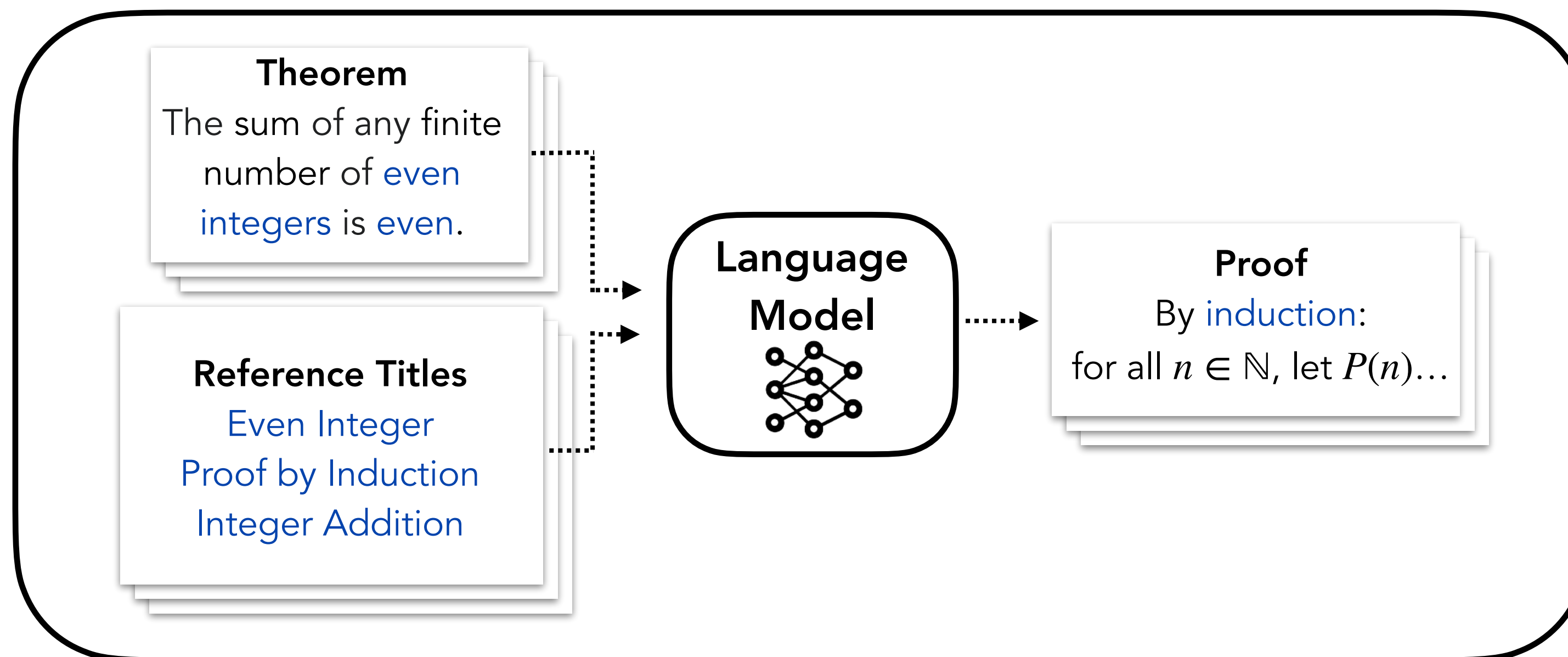


$$\max_{\theta} \log p_{\theta}(\mathbf{r}_{content} \mid \mathbf{r}_{title})$$

# NaturalProver

Condition on background knowledge (*references R*)  $p_{\theta}(\mathbf{y} \mid \mathbf{x}, R)$

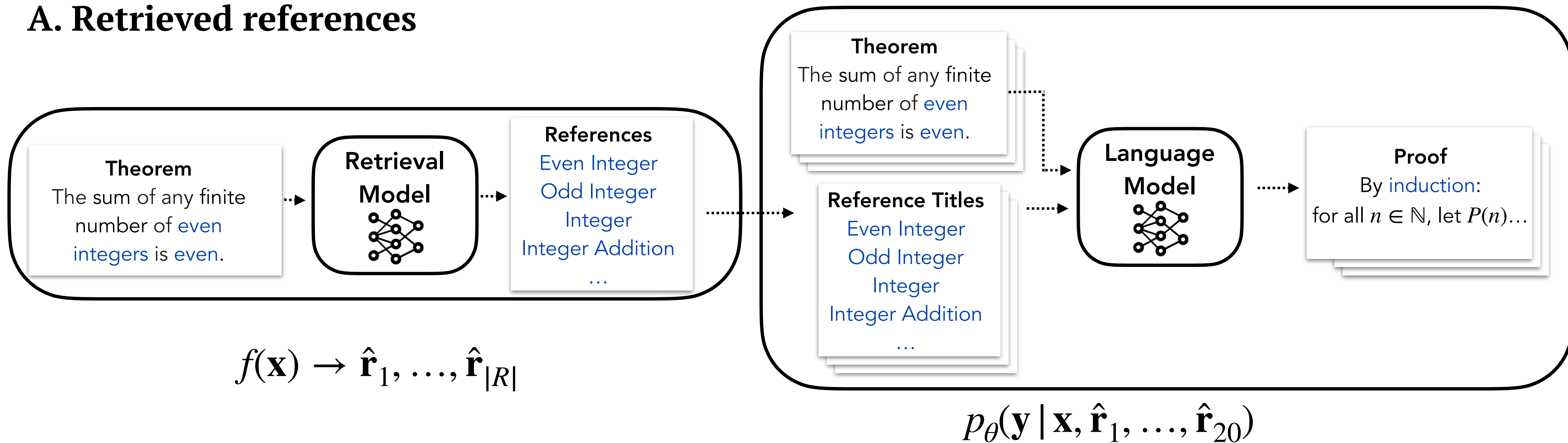
2. In-context reference titles:  $\max_{\theta} \log p_{\theta}(\mathbf{y} \mid \mathbf{x}, R_{title})$



# NaturalProver

Condition on background knowledge (*references R*)  $p_{\theta}(\mathbf{y} \mid \mathbf{x}, R)$

## A. Retrieved references

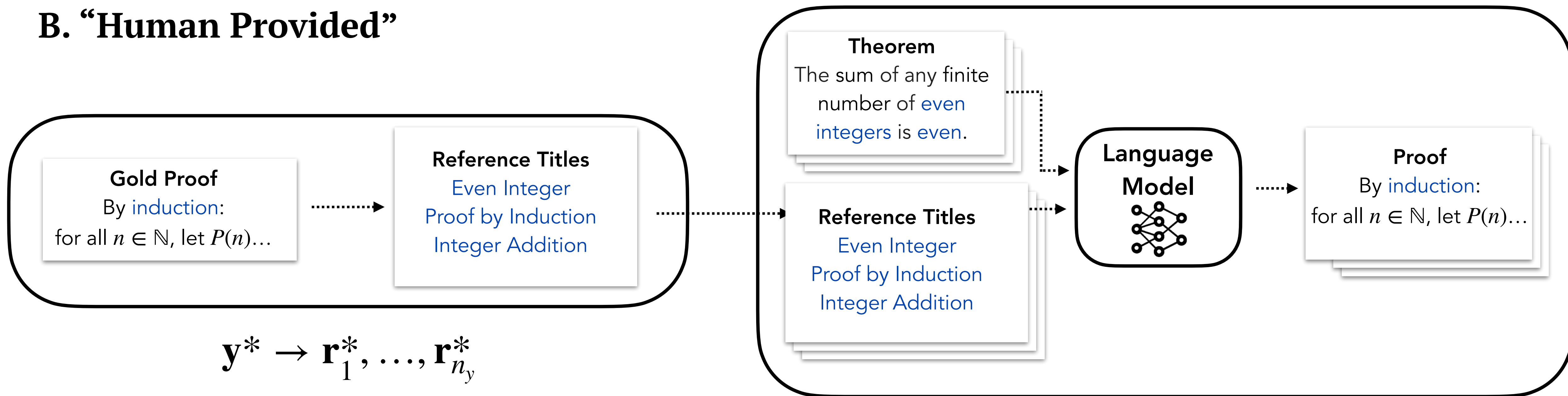




# NaturalProver

Condition on background knowledge (*references R*)

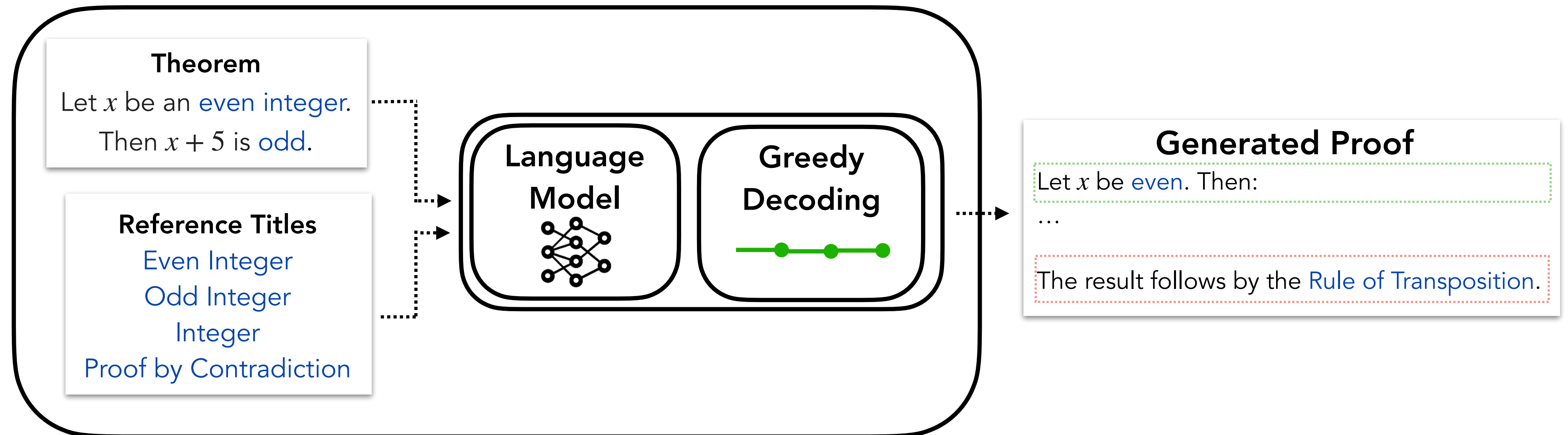
## B. “Human Provided”



$$p_{\theta}(y \mid \mathbf{x}, \mathbf{r}_1^*, \dots, \mathbf{r}_{n_y}^*)$$

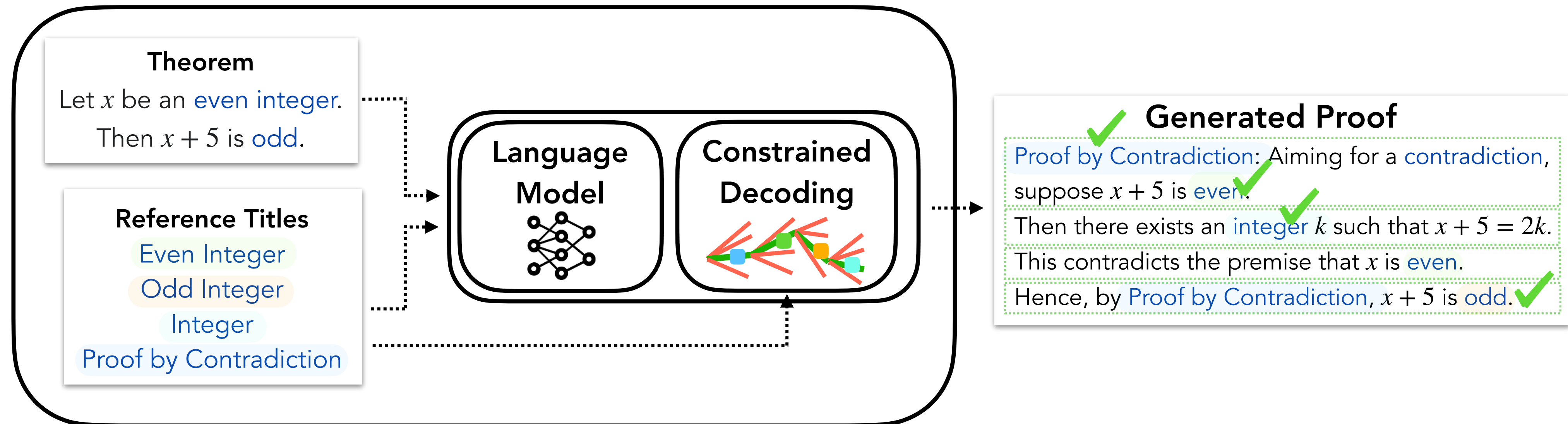
# NaturalProver + constrained decoding

## Generation



# NaturalProver + constrained decoding

## Generation



Hypothesis: searching for high-probability proofs that contain the provided references improves correctness

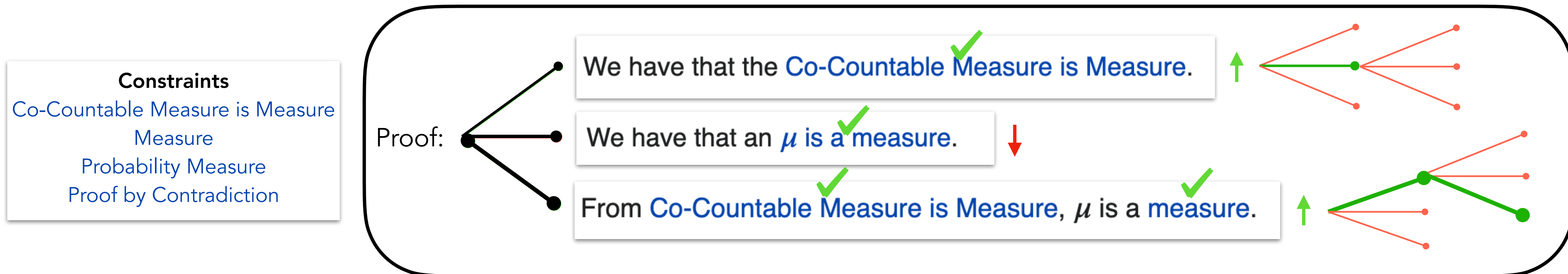
# NaturalProver + constrained decoding

## Stepwise beam search

- Sample multiple next-step candidates
- Retain those with high value in a beam
- Repeat

## Value function

$$v(y_{\leq t}) = \alpha \cdot \log p_{\theta}(y_{\leq t} | x) + \begin{matrix} \downarrow \uparrow \\ (1 - \alpha) \cdot \# \text{ constraints} \end{matrix} \quad \checkmark$$



# Method summary

## 1. Baseline $p_{\theta}(\mathbf{y} \mid \mathbf{x})$

## 2. NaturalProver-Retrieved

- $p_{\theta}(\mathbf{y} \mid \mathbf{x}, \hat{\mathbf{r}}_1, \dots, \hat{\mathbf{r}}_{20})$
- Top 20 retrievals from a NaturalProofs retriever  $f(\mathbf{x}) \rightarrow \hat{\mathbf{r}}_1, \dots, \hat{\mathbf{r}}_{|R|}$

## 3. NaturalProver

- $p_{\theta}(\mathbf{y} \mid \mathbf{x}, \mathbf{r}_1^*, \dots, \mathbf{r}_{n_y}^*)$
- Set of references from a ground-truth proof
- 4. **NaturalProver++** : with constrained decoding
- Language models: GPT-3 Curie
- Fine-tuned through OpenAI API

# Evaluation | Protocol

- Step-level annotation
  - $\{0,1\}$  errors (*stay tuned*)



- Main Page
- Community discussion
- Community portal
- Recent changes
- Random proof
- Glossary
- Help
- FAQ
- ProofWiki.org
- Proof Index
- Definition Index
- Symbol Index
- Axiom Index
- Mathematicians
- Books
- Sandbox
- All Categories
- Jokes
- To Do
- Proofread Articles
- Wanted Proofs
- More Wanted Proofs
- Stub Articles
- Tidy Articles
- Research Required
- Help Needed
- Questionable Content
- Improvements Invited
- Refactoring
- Missing Links
- Proposed Mergers
- Proposed Deletions
- Maintenance Needed
- Tools
- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Page information

## (EVALleo)/Even Integers form Commutative Ring/Thm17316/Proof0/Step2

### Contents [hide]

- 1 Theorem
- 2 Gold Proof
- 3 Prior Steps (Gold)
- 4 Current Step (Computer-Generated)
- 5 Questions

### Theorem

Let  $2\mathbb{Z}$  be the set of even integers.

Then  $(2\mathbb{Z}, +, \times)$  is a commutative ring.

However,  $(2\mathbb{Z}, +, \times)$  is not an integral domain.

### Gold Proof

From Integer Multiples form Commutative Ring,  $(2\mathbb{Z}, +, \times)$  is a commutative ring.

As  $2 \neq 1$ , we also have from Integer Multiples form Commutative Ring that  $(2\mathbb{Z}, +, \times)$  has no unity.

Hence by definition it is not an integral domain.



### Prior Steps (Gold)

From Integer Multiples form Commutative Ring,  $(2\mathbb{Z}, +, \times)$  is a commutative ring.

As  $2 \neq 1$ , we also have from Integer Multiples form Commutative Ring that  $(2\mathbb{Z}, +, \times)$  has no unity.

### Current Step (Computer-Generated)

Hence  $(2\mathbb{Z}, +, \times)$  is not an integral domain.

### Questions

#### Step Evaluation

1. Reasoning: Reference Aspects. Select all errors that appear in this step:

- Invalid Deployment: A statement deployed from a reference is not consistent with the reference.
- Invalid Justification: A reference is used as invalid justification for a statement.
- Non-existent Reference: A reference that does not exist is used.
- Self Loop: The step refers to the theorem itself.
- None of the above.

2. Reasoning: Equation Aspects. Select all errors that appear in this step:

- Invalid Equation: A standalone equation or initial equation in a derivation is invalid.
- Invalid Derivation: An equation in a derivation does not follow from the preceding steps.
- None of the above.

3. Reasoning: Other Aspects. Select all errors that appear in this step:

- Skips Steps: The step assumes unproven statements, or skips non-trivial steps.
- Repetition: The step repeats known things.
- Invalid (Other): The step's reasoning is invalid for reasons not captured by the other categories.



# Evaluation | Protocol

- Step-level annotation
  - $\{0,1\}$  errors
  - **Correctness**
  - **Usefulness**

7. *Correctness*. Is this step correct?

- **Yes**
- **No** (check this if you identified any error in previous questions)
- **Cannot determine** (e.g. this step makes a valid progress, but it depends on an invalid prior step)
- **This is a meaningless step** (e.g. QED)

---

8. *Helpfulness*. Could this step be a helpful hint for proving the theorem by myself?

- **Yes**
- **No**

# Evaluation | Protocol

- Full-proof summary
  - **Correctness**
  - **Usefulness**
- 15 [Applied] Mathematics students at UW (undergrad, master's, Ph.D.)
- 100 validation, 100 test theorems



Main Page  
Community discussion  
Community portal  
Recent changes  
Random proof  
Glossary  
Help  
FAQ

ProofWiki.org

Proof Index  
Definition Index  
Symbol Index  
Axiom Index  
Mathematicians  
Books  
Sandbox  
All Categories  
Jokes

To Do

Proofread Articles  
Wanted Proofs  
More Wanted Proofs  
Stub Articles  
Tidy Articles  
Research Required  
Help Needed  
Questionable Content  
Improvements Invited  
Refactoring  
Missing Links  
Proposed Mergers  
Proposed Deletions  
Maintenance Needed

Tools

What links here  
Related changes  
Special pages  
Printable version  
Permanent link  
Page information

Page [Discussion](#) [Read](#) [Edit](#) [View history](#)

## (EVALaries)/Even Integers form Commutative Ring/Thm17316/Proof0/Overall

**Contents** [hide]

- 1 Theorem
- 2 Gold Proof
- 3 Computer-Generated Proof
- 4 Questions

### Theorem

Let  $2\mathbb{Z}$  be the [set of even integers](#).

Then  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#).

However,  $(2\mathbb{Z}, +, \times)$  is not an [integral domain](#).

### Gold Proof

From [Integer Multiples form Commutative Ring](#),  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#).

As  $2 \neq 1$ , we also have from [Integer Multiples form Commutative Ring](#) that  $(2\mathbb{Z}, +, \times)$  has no unity.

Hence by definition it is not an [integral domain](#).

### Computer-Generated Proof

From [Even Integers form Commutative Ring with One](#), we have that  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#) with 1 element.

From [Ring of Integers Modulo Prime is Commutative Ring](#), we have that  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#) with no zero divisors.

However,  $(2\mathbb{Z}, +, \times)$  is not an [integral domain](#).

### Questions

#### Overall Evaluation

1. *Overall Score*. Choose a rating below. Not every statement in each rating will apply to the proof given the rating, but many statements will apply, and the general theme of the rating will hold:

- 0: The proof is missing.
- 1: The proof makes no sense or is unrelated to the problem statement.
- 2: The proof contains serious logical flaws and lacks adequate justification or explanation.
- 3: The proof has some gaps in reasoning.
- 4: The proof is correct or nearly correct and logically coherent.
- 5: The proof is correct and flows logically.

2. *Usefulness as an Aid*. Even if the proof is not perfect, would it be useful to you if you were to prove this theorem?

- 0: The proof is missing.
- 1: Seeing this proof would not help with proving the theorem by myself at all.
- 2: Seeing this proof would slightly decrease the effort needed to prove the theorem by myself.
- 3: Seeing this proof would make it substantially easier to prove the theorem by myself.
- 4: The proof is almost correct, and only needs a few minor corrections.
- 5: The proof is correct and could be directly used as a solution.



# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment

Invalid Justification

Hallucinated Ref.

Self Loop

## Prior Steps

---

Let  $x$  be an **even integer**.

## Current Step (Computer-Generated)

---

Then by definition of **even integer**:

$$\exists r \in \mathbb{N} : x = 2r$$

where  $r$  is not an **even integer**.

## Definition:Even Integer

An **integer**  $n \in \mathbb{Z}$  is **even** if and only if it is of the form:

$$n = 2r$$

where  $r \in \mathbb{Z}$  is an **integer**.

# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment

Invalid Justification

Hallucinated Ref.

Self Loop

## Prior Steps

$$\frac{1}{\cos \theta} = \sec \theta \quad \text{Secant is Reciprocal of Cosine}$$

## Current Step (Computer-Generated)

$$\rightsquigarrow \cos \theta = \frac{1}{\frac{1}{\sec \theta}} \quad \text{Cosine of Zero is One}$$

## Theorem: Cosine of Zero is One

$$\cos 0 = 1$$

where cos denotes the **cosine**.

# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment

Invalid Justification

Hallucinated Ref.

Self Loop

## Computer-Generated Proof

Let  $x \in S$ .

From **Clopen Points in Topological Space** we have that  $\{x\}$  is **open** in  $T$ .

From **Point is Open iff it is not Dense-in-itself**, it follows that  $\{x\}$  is not **dense-in-itself**.

# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment

Invalid Justification

Hallucinated Ref.

Self Loop

Theorem: Integral Domain is Reduced Ring

Let  $(D, +, \circ)$  be an *integral domain*.

Then  $D$  is *reduced*.

Computer-Generated Proof

By definition, an *integral domain* is a *ring*.

The result follows from *Integral Domain is Reduced Ring*.

# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment  
Invalid Justification  
Hallucinated Ref.  
Self Loop

---

## Reasoning: Equation

Invalid Equation  
Invalid Derivation

---

## Current Step (Computer-Generated)

---

$$\frac{1}{1-x} = \frac{1}{\frac{1}{2} + \frac{1}{2(1-x)}}$$

Reciprocal of One plus x by Two plus x fourth plus x fifth

# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment  
Invalid Justification  
Hallucinated Ref.  
Self Loop

---

## Reasoning: Equation

Invalid Equation  
Invalid Derivation

---

Then:

$$x = 2n - 5$$

Current Step (Computer-Generated)

---

$$= 2(n + 1) - 5$$



# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment  
Invalid Justification  
Hallucinated Ref.  
Self Loop

---

## Reasoning: Equation

Invalid Equation  
Invalid Derivation

---

## Reasoning: Other

Skips Steps  
Repetition  
Invalid (Other)

### Computer-Generated Proof

---

We have that a metric space is perfectly  $T_4$ .

Then we have that a metric space is perfectly normal.

■

By definition, a topological space is perfectly normal space if and only if it is:

a perfectly  $T_4$  space  
a  $T_1$  (Fréchet) space.

# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment  
Invalid Justification  
Hallucinated Ref.  
Self Loop

## Reasoning: Equation

Invalid Equation  
Invalid Derivation

## Reasoning: Other

Skips Steps  
Repetition  
Invalid (Other)

### Computer-Generated Proof

#### Necessary Condition

Let  $f$  be **continuous** on  $\mathbb{I}$ .

Let  $x \in \mathbb{I}$ .

Then by definition of **continuity** there exists  $\delta > 0$  such that  $0 < |x - x| < \delta$ .

Then:

$$\begin{aligned} |f(x) - f(x)| &= |f'(x) - f'(x)| && \text{where } f' \text{ is the derivative of } f \\ &= |f'(x)| && \text{Derivative of Constant Multiple} \\ &< \delta \\ &= \epsilon && \text{where } \epsilon > 0 \end{aligned}$$

□

#### Sufficient Condition

Let  $f$  satisfy the condition given.

Let  $x \in \mathbb{I}$ .

Then by definition of **continuity** there exists  $\delta > 0$  such that  $0 < |x - x| < \delta$ .

Then:

$$\begin{aligned} |f(x) - f(x)| &= |f'(x) - f'(x)| && \text{where } f' \text{ is the derivative of } f \\ &= |f'(x)| && \text{Derivative of Constant Multiple} \\ &< \delta \\ &= \epsilon && \text{where } \epsilon > 0 \end{aligned}$$

Hence the result.

Actual prediction from baseline GPT-3 model



# Evaluation | Schema

## Reasoning: Reference

Invalid Deployment  
Invalid Justification  
Hallucinated Ref.  
Self Loop

---

## Reasoning: Equation

Invalid Equation  
Invalid Derivation

---

## Reasoning: Other

Skips Steps  
Repetition  
Invalid (Other)

---

Language

---

Symbolic

## Current Step (Computer-Generated)

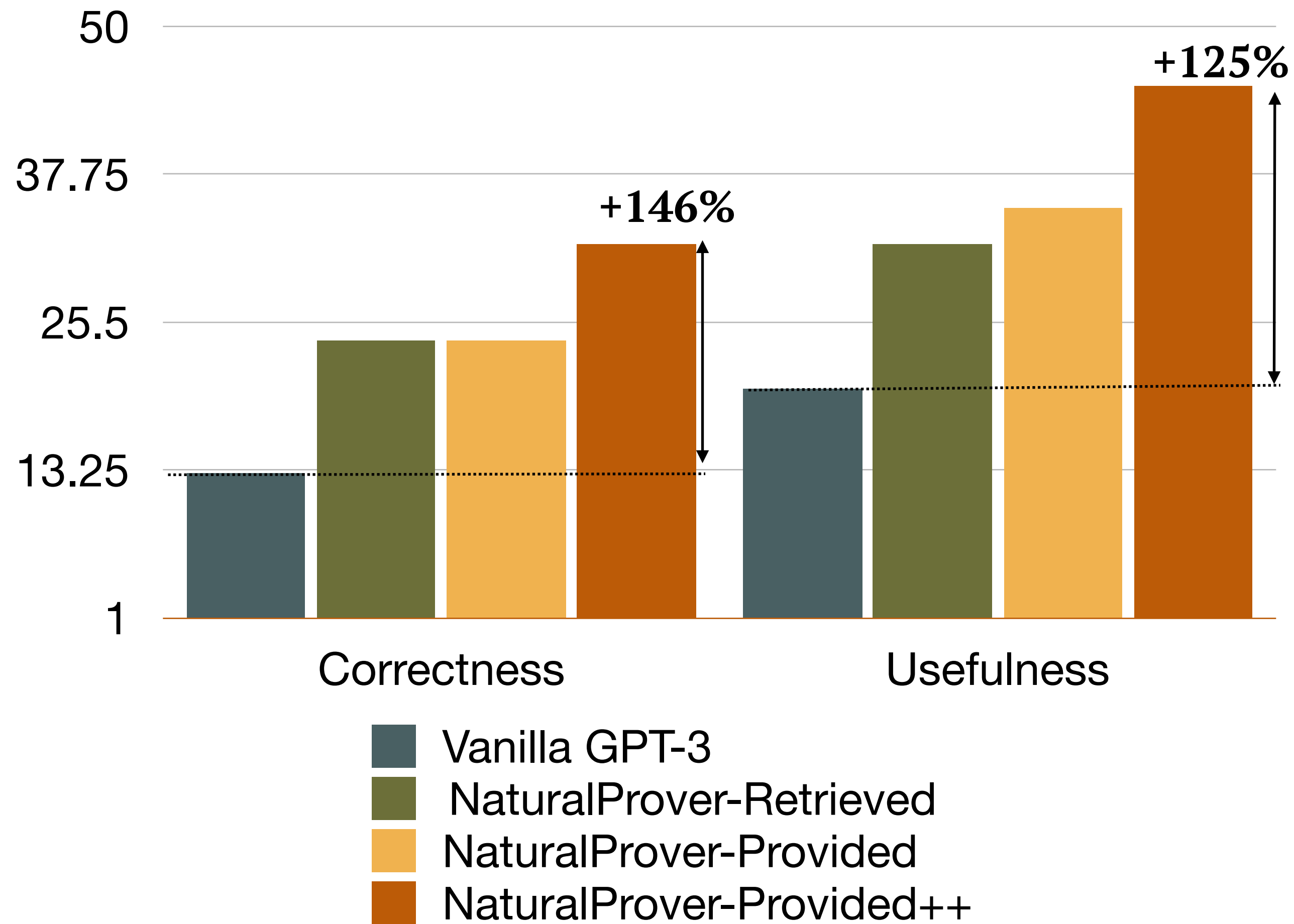
---

$$\mathbf{a} \times 0 = \backslash\text{zero}$$

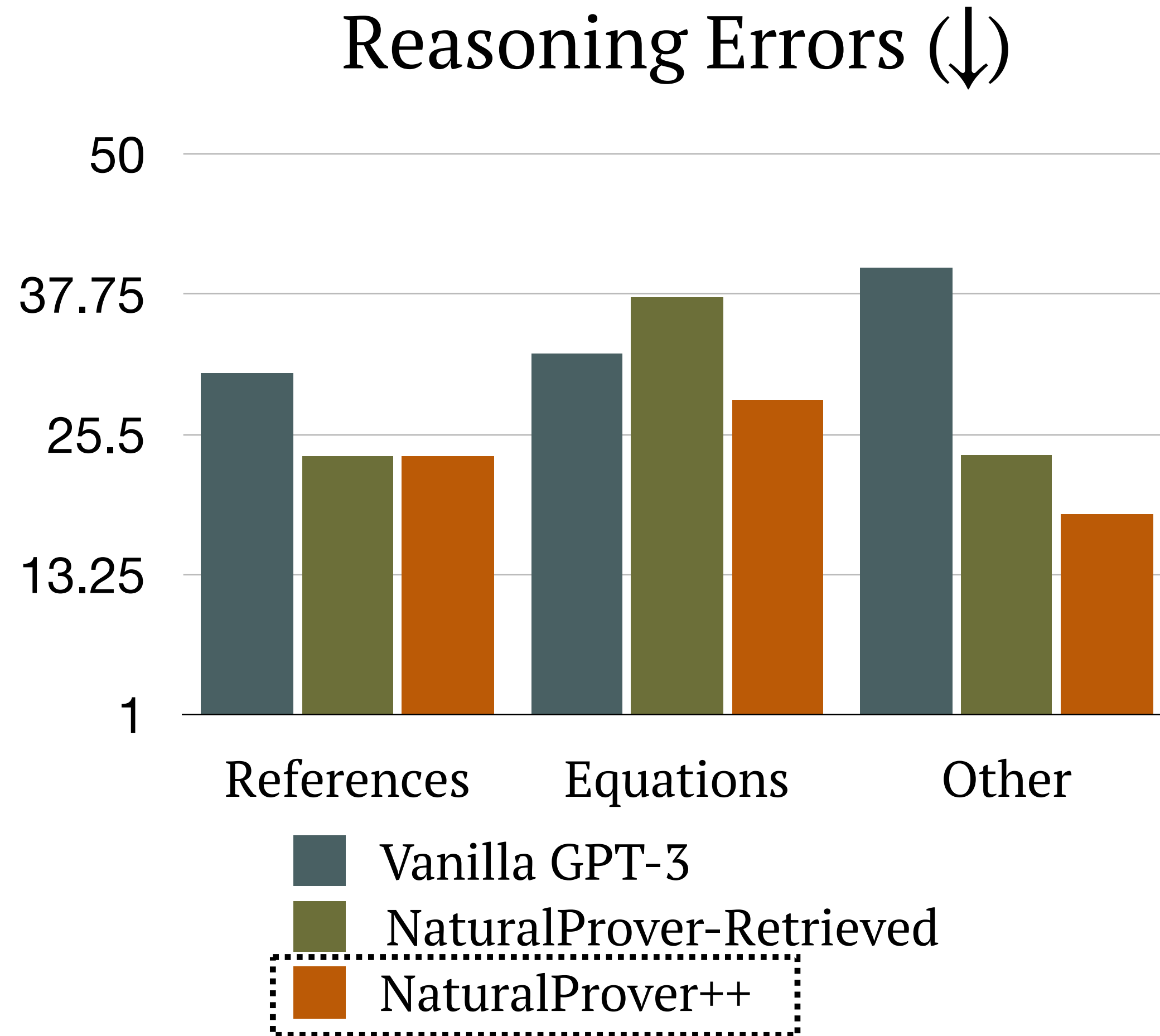
Definition of [Scalar Multiplication on Vector Quantity](#)

# Results | Full Proof Summary

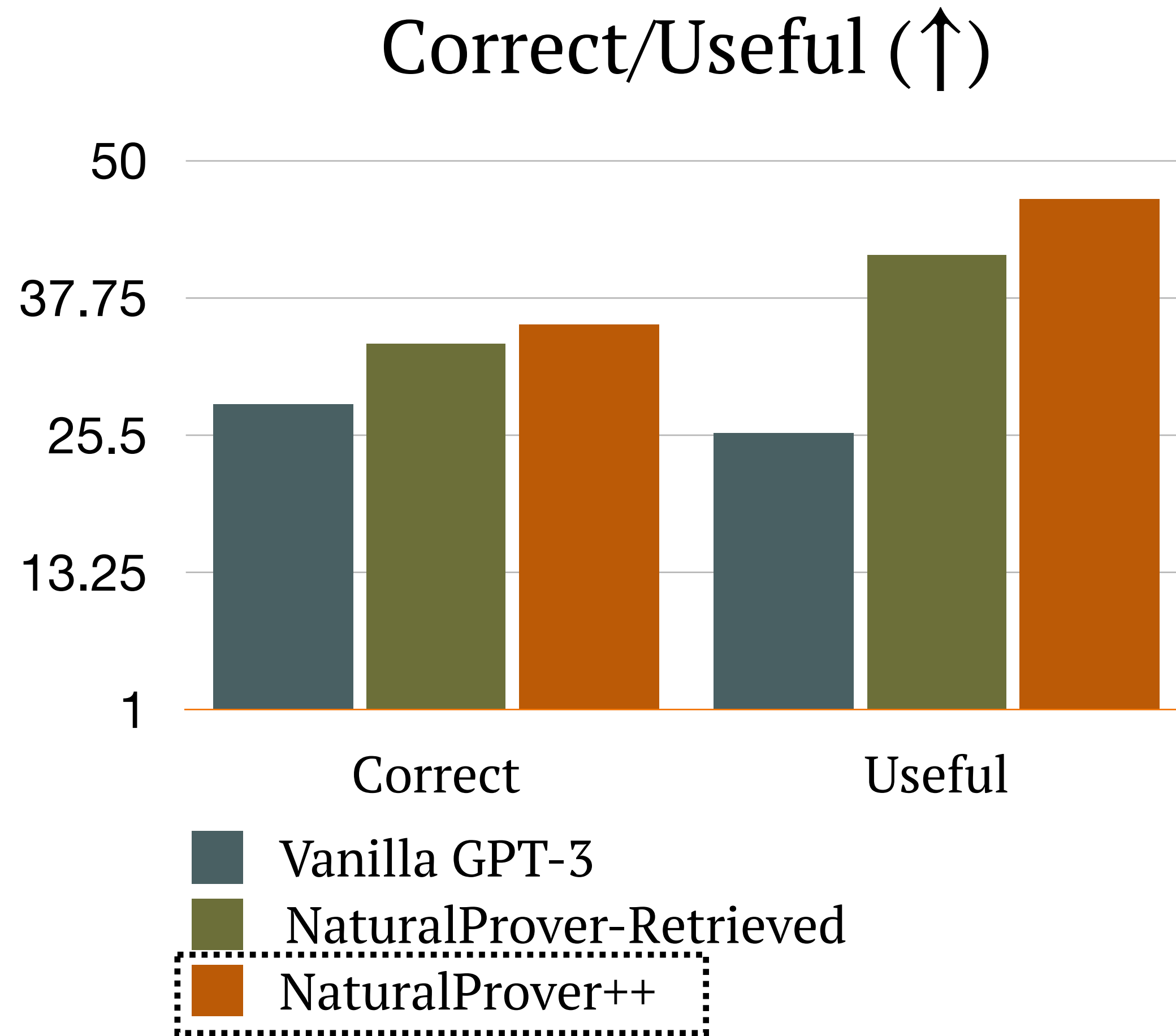
## Full Proof Summary Metrics



# Results | Per-step Metrics



# Results | Per-step Metrics



# Type 1: "Reference assembly"

## Theorem: Compact Complement Topology is Connected

Let  $T = (\mathbb{R}, \tau)$  be the compact complement topology.

Then  $T$  is a connected space.

### Retrieved References

Irreducible Space is Connected  
Compact Complement Topology is Irreducible  
Definition: Compact Space/Topology/Subspace  
Definition: Connected (Topology)/Topological Space  
Definition: Open Set/Topology  
Path-Connected Space is Connected  
Irreducible Space is Locally Connected  
Definition: Separation (Topology)  
Definition: Euclidean Space/Euclidean Topology/Real Number Line  
Definition: Bounded Above Set  
Definition: Open Cover  
Definition: Compact Complement Topology  
Finite Complement Space is Irreducible  
Definition: Irreducible Space  
Compact Complement Topology is Compact  
Definition: Finite Set  
Finite Complement Space is Locally Connected  
Definition: Clopen Set  
Definition: Disjoint Sets  
Countable Complement Space is Irreducible

### Proof: NaturalProver-Retrieved

Follows from:

Compact Complement Topology is Irreducible

Irreducible Space is Connected



Uses theorems

Correct & Useful ✓

## Theorem: Compact Complement Topology is Connected

Let  $T = (\mathbb{R}, \tau)$  be the compact complement topology.

Then  $T$  is a connected space.

### Proof: Baseline GPT-3

Let  $U \in \tau$  be any open set of  $T$ .

Then from **Basis for Compact Complement Topology**, the set:

$$\mathcal{B} := \{(n-1, n) : n \in \mathbb{N}_{>0}\}$$

is a basis for  $T$ .

That is, every open set of  $T$  contains a compact subset which is open.

So every open set of  $T$  contains a compact subset which is open.

So by definition,  $T$  is connected.

#### Reasoning: Reference

Invalid Deployment

Invalid Justification

Hallucinated Ref.

Self Loop

#### Reasoning: Equation

Invalid Equation

Invalid Derivation

#### Reasoning: Other

Skips Steps

Repetition

Invalid (Other)

#### Language

Symbolic



## Type 2: "Template Adaptation"

### Theorem: Pointwise Addition on Real-Valued Functions is Associative

---

Let  $f, g, h : S \rightarrow \mathbb{R}$  be **real-valued functions**.

Let  $f + g : S \rightarrow \mathbb{R}$  denote the **pointwise sum of  $f$  and  $g$** .

Then:

$$(f + g) + h = f + (g + h)$$

### Proof: NaturalProver++

---

$$\forall x \in S : ((f + g) + h)(x) = (f(x) + g(x)) + h(x)$$

Definition of **Pointwise Addition of Real-Valued Functions**



$$= f(x) + (g(x) + h(x))$$

**Real Addition is Associative**



$$= (f + (g + h))(x)$$

Definition of **Pointwise Addition of Real-Valued Functions**



# Type 2: "Template Adaptation"

Theorem: Pointwise Addition of Real-Valued Functions is Associative

Let  $f, g, h : S \rightarrow \mathbb{R}$  be real-valued functions.

Let  $f + g : S \rightarrow \mathbb{R}$  denote the pointwise sum of  $f$  and  $g$ .

Then:

$$(f + g) + h = f + (g + h)$$

## Pointwise Addition on Complex-Valued Functions is Associative

### Theorem

Let  $S$  be a set.

Let  $f, g, h : S \rightarrow \mathbb{C}$  be complex-valued functions.

Let  $f + g : S \rightarrow \mathbb{C}$  denote the pointwise sum of  $f$  and  $g$ .

Then:

$$(f + g) + h = f + (g + h)$$

That is, pointwise addition on complex-valued functions is associative.

### Proof

$$\begin{aligned} \forall x \in S : ((f + g) + h)(x) &= (f(x) + g(x)) + h(x) && \text{Definition of Pointwise Addition of Complex-Valued Functions} \\ &= f(x) + (g(x) + h(x)) && \text{Complex Addition is Associative} \\ &= (f + (g + h))(x) && \text{Definition of Pointwise Addition of Complex-Valued Functions} \end{aligned}$$

■

## Proof: NaturalProver++

$$\begin{aligned} \forall x \in S : ((f + g) + h)(x) &= (f(x) + g(x)) + h(x) \\ &= f(x) + (g(x) + h(x)) \\ &= (f + (g + h))(x) \end{aligned}$$

Definition of Pointwise Addition of Real-Valued Functions  
Real Addition is Associative  
Definition of Pointwise Addition of Real-Valued Functions



# Type 3: "sophisticated"

## Theorem: Singleton Set is not Dense-in-Itself

Let  $T = (S, \tau)$  be a topological space.

Let  $x \in S$ .

Then the singleton set  $\{x\}$  is not dense-in-itself.

## Proof: NaturalProver++

From **Singleton Point is Isolated**,  $\{x\}$  has an isolated point.

Hence the result by definition of **dense-in-itself**.

Uses definitions

Uses theorems

Correct & Useful ✓

### Theorem

Let  $T = (S, \tau)$  be a topological space.

Let  $x \in S$ .

Then  $x$  is an isolated point of the singleton set  $\{x\}$ , but not necessarily an isolated point of  $T$ .

### Definition

Let  $T = (S, \tau)$  be a topological space.

Let  $H \subseteq S$ .

Then  $H$  is **dense-in-itself** if and only if it contains no isolated points.

# Type 3: "sophisticated"

## Theorem: Singleton Set is not Dense-in-Itself

---

Let  $T = (S, \tau)$  be a [topological space](#).

Let  $x \in S$ .

Then the [singleton set](#)  $\{x\}$  is not [dense-in-itself](#).

### Reasoning: Reference

Invalid Deployment

Invalid Justification

[Hallucinated Ref.](#)

Self Loop

## Proof: Baseline GPT-3

---

Let  $x \in S$ .

From [Clopen Points in Topological Space](#) we have that  $\{x\}$  is [open](#) in  $T$ .

From [Point is Open iff it is not Dense-in-itself](#), it follows that  $\{x\}$  is not [dense-in-itself](#).



# Results | Per-step Metrics

	Lexical		Grounding				Halluc (↓)
	GLEU	Token F1	kF1	Ref-P	Ref-R	Ref-F1	
GPT-3	24.40	49.96	49.30	29.93	24.73	23.69	17.92
NATURALPROVER <sub>RETRIEVE</sub>	26.58	53.02	55.88	38.17	28.48	27.10	2.25
NATURALPROVER	35.27	66.00	90.07	93.05	86.05	87.08	1.60
NATURALPROVER <sub>++</sub>	34.49	65.61	96.39	94.66	95.00	93.92	1.71

human metric	Correctness [full]	0.93	0.91	0.86	0.83	0.85	0.85	0.94
	Usefulness [full]	0.90	0.87	0.82	0.78	0.80	0.80	0.97
	Correctness [step]	0.81	0.80	0.74	0.69	0.73	0.72	0.97
	Usefulness [step]	0.65	0.61	0.53	0.47	0.52	0.51	0.98
	Reasoning Errors: Ref.	0.71	0.64	0.52	0.48	0.50	0.50	0.95
	Reasoning Errors: Eqn.	0.70	0.74	0.75	0.69	0.74	0.73	0.78
	Reasoning Errors: Other	0.65	0.61	0.53	0.47	0.52	0.51	0.98
	Language Errors	0.99	1.00	0.99	0.98	0.99	0.99	0.73
Symbolic Errors	-0.72	-0.80	-0.88	-0.89	-0.89	-0.88	-0.21	

Pearson correlation between human and automatic metrics

# Next-step suggestion

Theorem  $\mathbf{x}$

Even Integers form Commutative Ring

---

Let  $2\mathbb{Z}$  be the [set of even integers](#).

Then  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#).

However,  $(2\mathbb{Z}, +, \times)$  is not an [integral domain](#).

**Input**

Prior proof steps  $y_{<t}^*$

Human-written

Prior Steps

---

From [Integer Multiples form Commutative Ring](#),  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#).

As  $2 \neq 1$ , we also have from [Integer Multiples form Commutative Ring](#) that  $(2\mathbb{Z}, +, \times)$  has no [unity](#).

Next-step suggestions  
 $\{y_t^k\}_{k=1}^K$

Thus  $(2\mathbb{Z}, +, \times)$  is not an [integral domain](#).

---

So, by definition,  $(2\mathbb{Z}, +, \times)$  is not an [integral domain](#).

---

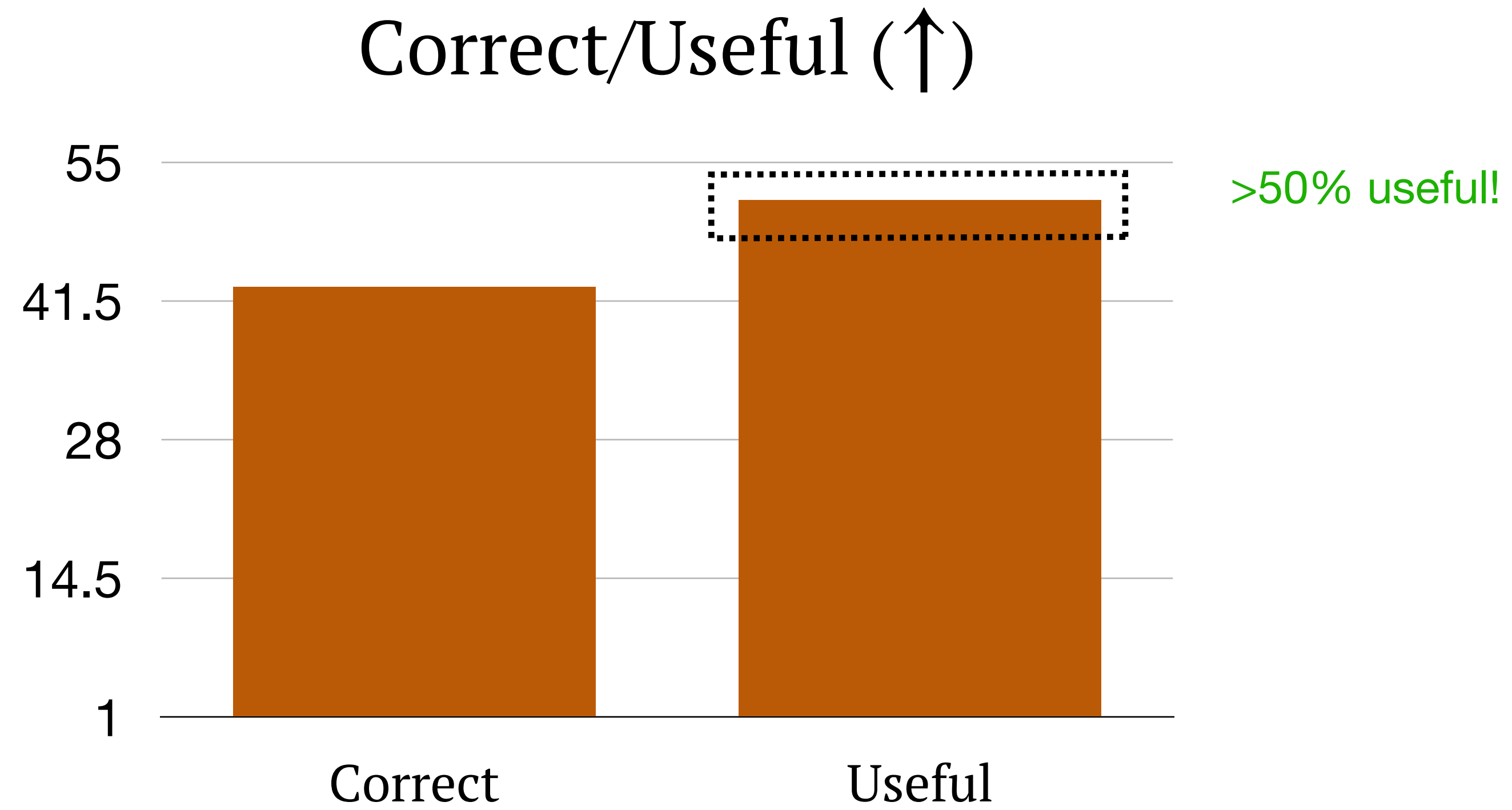
So, while  $(2\mathbb{Z}, +, \times)$  is a [commutative ring](#), it is not an [integral domain](#).

---

Hence  $(2\mathbb{Z}, +, \times)$  is not an [integral domain](#).



**Outputs**

# Next-step suggestion





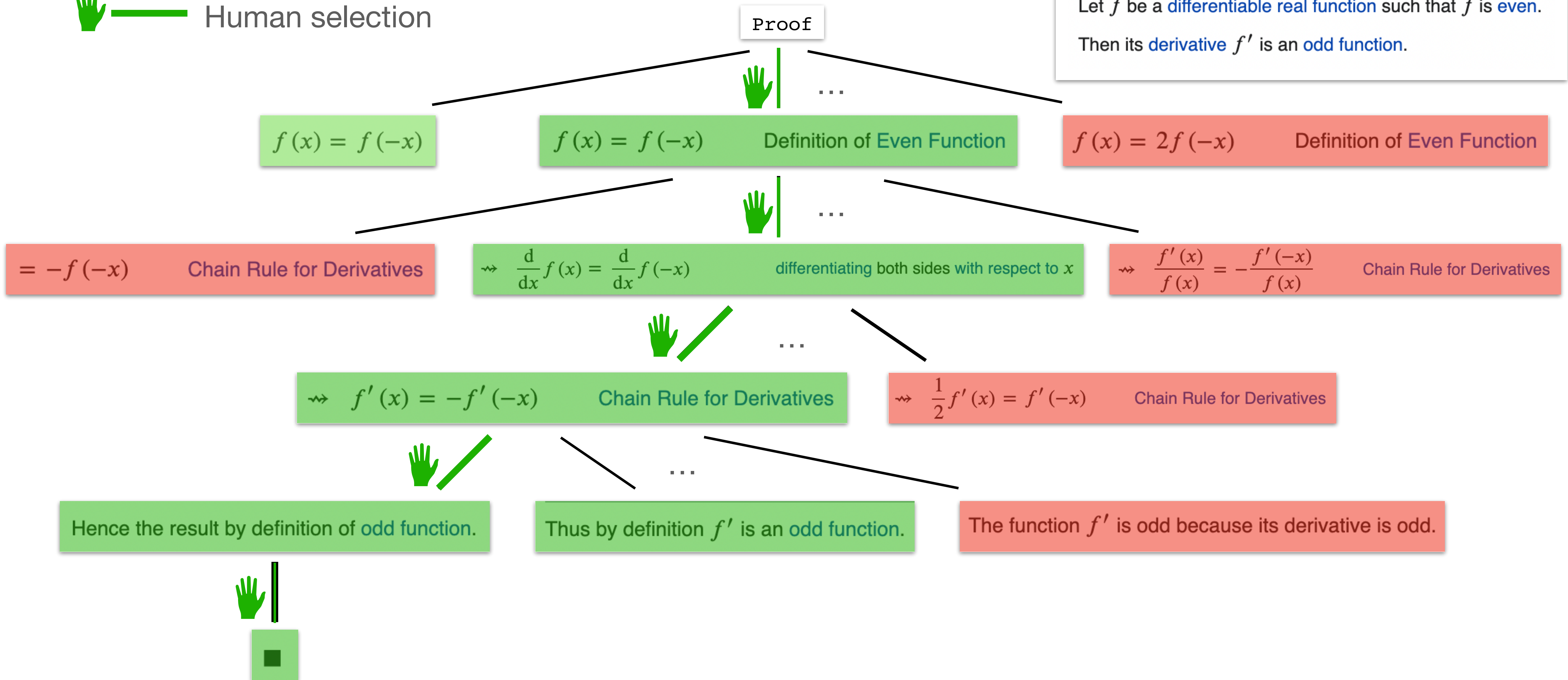
# Human-machine collaborative proving

 Generated suggestion  
 Human selection

Derivative of Even Function is Odd

---

Let  $f$  be a differentiable real function such that  $f$  is even.  
Then its derivative  $f'$  is an odd function.



# Human-machine collaborative proving

Derivative of Even Function is Odd

Let  $f$  be a **differentiable real function** such that  $f$  is **even**.

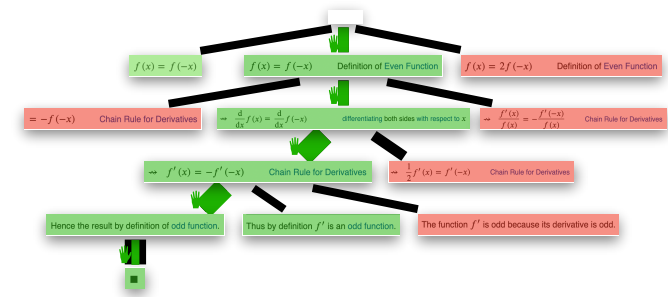
Then its **derivative**  $f'$  is an **odd function**.

$f(x) = f(-x)$       Definition of **Even Function**

$\rightsquigarrow \frac{d}{dx} f(x) = \frac{d}{dx} f(-x)$       **differentiating both sides with respect to  $x$**

$\rightsquigarrow f'(x) = -f'(-x)$       **Chain Rule for Derivatives**

Hence the result by definition of **odd function**.



# Where do models struggle the most?

I: Equations / Derivations

> 25% error rate for all models

Let  $x \in \mathbb{Z}$  be **even**.

Then:

$$x + 5 = 7$$

$\rightsquigarrow$   $x = 4 + \frac{1}{2}$   
 $= 4 + \frac{1 \pm 1}{2}$       Definition of **Odd Integer**

and so  $x$  is **odd**.

### Computer-Generated Proof

From the definition of the **discrete uniform distribution**,  $X$  has **probability mass function**:

$$\Pr(X = k) = \frac{n}{n + k}$$

From the definition of a **moment generating function**:

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=0}^n \Pr(X = k) e^{tk}$$

So:

$$M_X(t) = \sum_{k=0}^n \Pr(X = k) e^{tk}$$
$$= \frac{n}{n+k} \sum_{k=0}^n \left( (e^{tk})^n \prod_{j=0}^{n-1} (e^{t \frac{j+1}{n}}) \right)$$
      Sum of Geometric Sequence
$$= \frac{n}{n+k} \sum_{k=0}^n (e^{t \frac{j+1}{n}} e^{tj})$$
$$= \frac{n}{n+k} (1 - e^{nt})$$
      Definition of **Discrete Uniform Distribution**
$$= \frac{e^t (1 - e^{nt})}{n(1 - e^t)}$$

# Where do models struggle the most?

II: Reference usage

> 15% error rate for all models

---

<b>Theorem 1</b>	<b>Geometric Mean of two Positive Real Numbers is Between them</b> Let $a, b \in \mathbb{R}$ be <u>real numbers</u> such that $0 < a < b$ . Let $G(a, b)$ denote the <u>geometric mean</u> of $a$ and $b$ . Then: $a < G(a, b) < b$	
NATURALPROVER <sub>++</sub>	<b>Proof:</b> By definition of <u>geometric mean</u> : $G(a, b) = \frac{a^2 + b^2}{2}$	correct <input checked="" type="checkbox"/> deploy <input checked="" type="checkbox"/>
.....		
<i>Gold Proof</i>	<b>Proof:</b> By definition of <u>geometric mean</u> : $G(a, b) := \sqrt{ab}$ where $\sqrt{ab}$ specifically denotes the <u>positive square root</u> of $a$ and $b$ .	
.....		
<i>Reference</i>	<u>Geometric Mean</u> Let $x_1, x_2, \dots, x_n \in \mathbb{R}_{>0}$ be <u>(strictly) positive real numbers</u> . The geometric mean of $x_1, x_2, \dots, x_n$ is defined as: $G_n := (\prod_{k=1}^n x_k)^{1/n}.$	

# Where do models struggle the most?

## III: Long proofs

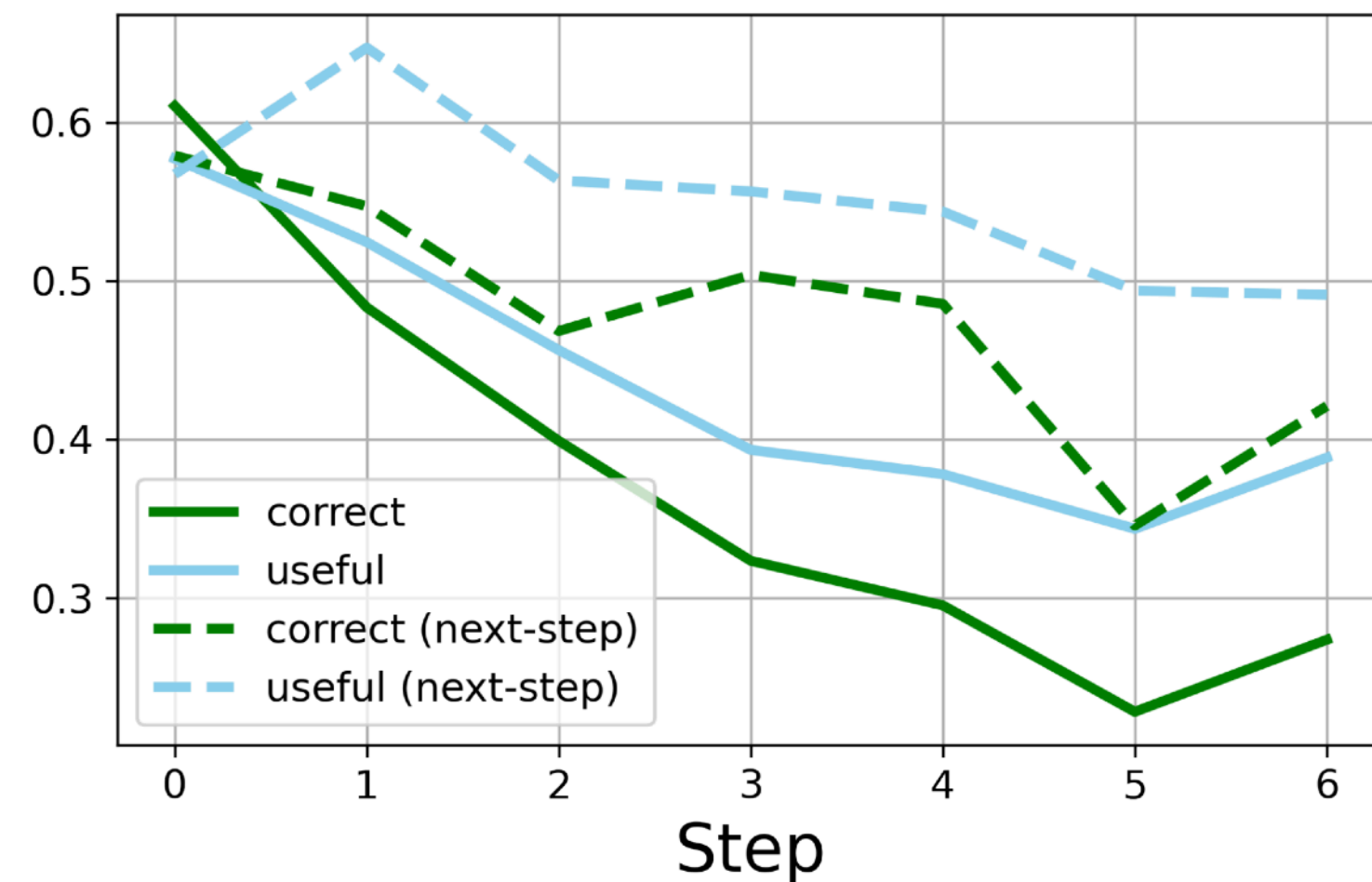


Figure 2: Per-step correctness and usefulness as a function of step number, for full-proof generation with `NATURALPROVER++` and next-step prediction with `NATURALPROVER`.



# Looking ahead | reliability

**Proof:**

$$0 < 1$$

$$\rightsquigarrow -0 > -1$$

$$\rightsquigarrow 0 > -1$$

$$\rightsquigarrow -1 < 0$$

■

Real Zero is Less than Real One

Order of Real Numbers is Dual of Order of their Negatives

Negative of Real Zero equals Zero

Dual Ordering

**Proof:**

$$0 < 1$$

$$\rightsquigarrow -1 < 0$$

$$\rightsquigarrow -1 < 0$$

■

Real Zero is Less than Real One

Order of Real Numbers is Dual of Order of their Negatives

Negative of Real Zero equals Zero

Detectable?



# Looking ahead | capability

**Proof:**

$$0 < 1$$

$$\rightsquigarrow -1 < 0$$

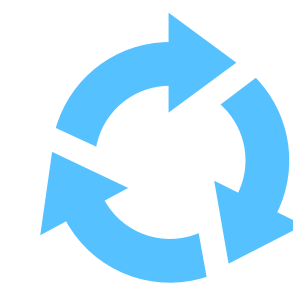
$$\rightsquigarrow -1 < 0$$



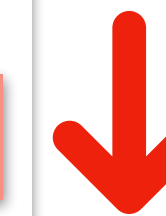
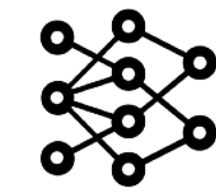
Real Zero is Less than Real One

Order of Real Numbers is Dual of Order of their Negatives

Negative of Real Zero equals Zero



Language model



Proof

$$f(x) = f(-x)$$



$$f(x) = f(-x)$$

Definition of Even Function



$$f(x) = 2f(-x)$$

Definition of Even Function

Learning from feedback

# Looking ahead | informal $\leftrightarrow$ formal



## Theorem

Let  $f : S \rightarrow \mathbb{R}$  be a **real function** on some **subset**  $S$  of  $\mathbb{R}$ .

Let  $I \subseteq S$  be a **real interval**.

Let  $f : I \rightarrow \mathbb{R}$  be **continuous** on  $I$ .

Then  $f$  is a **Darboux function**.

That is:

Let  $a, b \in I$ .

Let  $k \in \mathbb{R}$  lie between  $f(a)$  and  $f(b)$ .

That is, either:

$$f(a) < k < f(b)$$

or:

$$f(b) < k < f(a)$$

Then  $\exists c \in (a..b)$  such that  $f(c) = k$ .

```
/-- **Intermediate Value Theorem** for continuous functions on c
`f a ≤ t ≤ f b`.-/
lemma intermediate_value_Icc {a b : α} (hab : a ≤ b) {f : α → δ}
  Icc (f a) (f b) ⊆ f '' (Icc a b) :=
  is_preconnected_Icc.intermediate_value (left_mem_Icc.2 hab) (right)
```

```
subsection <Intermediate Value Theorem>

lemma IVT':
  fixes f :: "'a::linear_continuum_topology ⇒ 'b::linorder_topology"
  assumes y: "f a ≤ y" "y ≤ f b" "a ≤ b"
  and *: "continuous_on {a..b} f"
  shows "∃x. a ≤ x ∧ x ≤ b ∧ f x = y"
proof -
  have "connected {a..b}"
  unfolding connected_iff_interval by auto
  from connected_continuous_image[OF * this, THEN connectedD_interval, of "f a" "f b" y] y
  show ?thesis
  by (auto simp add: atLeastAtMost_def atLeast_def atMost_def)
qed
```

Best of both worlds?

- Flexibility
- Reliability

- Flexibility
- Reliability

# Thank you!

- Full paper : <https://arxiv.org/pdf/2205.12910.pdf>
- Code/data/models : <https://github.com/wellecks/naturalprover>

## NaturalProver: Grounded Mathematical Proof Generation with Language Models

[NaturalProver: Grounded Mathematical Proof Generation with Language Models](#)  
Sean Welleck\*, Jiacheng Liu\*, Ximing Lu, Hannaneh Hajishirzi, Yejin Choi

This repo contains:

- The **NaturalProofs-Gen** datasets.
- **GPT-3, GPT-J, GPT-2** code for training and generation.
- **Automatic evaluation** for proof generation and next-step suggestion.
- **GPT-2** trained model.