Relationalizing Tables with LLMs: The Promise and Challenges

Zachary Huang, Eugene Wu
Columbia University

Outline

1. **Problem of relationalization, and the SOTA solution**

2. **Better LLMs and prompts greatly improve accuracy**

   With GPT-4, CoT, and Decomposed Prompt, acc 57.0% → 74.6%

3. **Error analysis reveals the ambiguity challenge**

   Of the 25.4% errors, ~half are due to ambiguity

Problem: Tables in the wild are not relational

COVID-19 Data

Date values as column headers

Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20
Afghanistan	33.93911	67.709953	0	0	0	0	0	0
Albania	41.1533	20.1683	0	0	0	0	0	0
Algeria	28.0339	1.6596	0	0	0	0	0	0
Andorra	42.5063	1.5218	0	0	0	0	0	0
Angola	-11.2027	17.8739	0	0	0	0	0	0
Antigua and Barbuda	17.0608	-61.7964	0	0	0	0	0	0
Argentina	-38.4161	-63.6167	0	0	0	0	0	0
Armenia	40.0691	45.0382	0	0	0	0	0	0
Australia	-35.4735	149.0124	0	0	0	0	0	0

Problem: Tables in the wild are not relational

Netflix Kaggle Dataset

Netflix Movies and TV Shows

Data Card | Code (21) | Discussion (1) | Suggestions (0)

›r TV	title Title of the Movie / Tv Show	director Director of the Movie	cast Actors involved in the movie / show
	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Hamish Linklater, Henry Thomas, Kristin Lehman, Samantha Sloyan, Igby Rig...
	My Little Pony: A New Generation	Robert Cullen, José Luis Ucha	Vanessa Hudgens, Kimiko Glenn, James Marsden, Sofia Carson, Liza Koshy, Ken Jeong, Elizabeth Perkins...
	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick Medley, Mutabaruka, Afemo Omilami, Reggie Ca...

List of actors in a single cell (against 1NF)

To relationalize tables, previous work...

Designed 7 operators

1 Pivot

Name	Alice
Age	17
Name	Bob
Age	19

Name	Age
Alice	17
Bob	19

2 Transpose

Name	Alice	Bob
Age	17	19

Name	Age
Alice	17
Bob	19

3 Subtitle

Name	Age
Student	
Alice	17
Bob	19
Teacher	
Claire	32

Name	Age	Role
Alice	17	Student
Bob	19	Student
Claire	32	Teacher

4 Ffill

Role	Name
Student	Alice
	Bob
Teacher	Claire
	David

Role	Name
Student	Alice
Student	Bob
Teacher	Claire
Teacher	David

Li, Peng, et al. "Auto-Tables: Synthesizing Multi-Step Transformations to Relationalize Tables without Using Examples." *VLDB* 2023

To relationalize tables, previous work...

Designed 7 operators

5 Explode

Teacher	Student
Claire	Alice, Bob
David	Eva, Fiona

→

Teacher	Student
Claire	Alice
Claire	Bob
David	Eva
David	Fiona

6 Stack

Name	2013	2014	2015
Alice	A	B	A

→

Name	Year	Grade
Alice	2013	A
Alice	2014	B
Alice	2015	A

7 Wide_to_long

Name	2013 English	2014 English	2015 English	2013 Math	2014 Math	2015 Math
Alice	A	B	A	B	B	A

→

Name	Year	Course	Grade
Alice	2013	English	A
Alice	2014	English	B
Alice	2015	English	A
Alice	2013	Math	B
Alice	2014	Math	B
Alice	2015	Math	A

Li, Peng, et al. "Auto-Tables: Synthesizing Multi-Step Transformations to Relationalize Tables without Using Examples." *VLDB* 2023

To relationalize tables, previous work...

Given a table, to automatically find the operators

Benchmark: Real-world datasets collected by Li et al.(2023)

Method	Acc (%)
Auto-Tables [Li et al.(2023)]	57.0
GPT-3.5	13.1

Specialized DNN

Few-shot
in-context learning

To relationalize tables, previous work...

Given a table, to automatically find the operators

Benchmark: Real-world datasets collected by Li et al.(2023)

Method	Acc (%)
Auto-Tables [Li et al.(2023)]	57.0
GPT-3.5	13.1

Specialized DNN

Few-shot
in-context learning

Better prompts/models can help?

To relationalize tables, previous work...

How much do better prompts/models help?

Describe 7 operators

No Chain of Thought (CoT) What if CoT?

Examples for each operator for few-shot learning

Task: Predict transformation operators for table.
=========
Operator descriptions:
- Stack: collapse homogeneous cols into rows.
@param (int): start_idx: zero-based starting column index...
=========
Output: JSON. E.g., [{"operator": "Ffill", "end_idx": 1}].
No explanation is needed.
=========
Example inputs and outputs
Input
|id|date|items|
|1|2021-01-01|apple, banana, orange|
Output
[{"operator": "Explode", "column_idx": 2}]...
=========
Your task:
Input: ...
Output:

Large prompt with too many works! What if decompose?

GPT-3.5! What if -4?

Task Decomposition

Exploit the sequence of operators

1. Structural Change of **All Rows and Columns** (*Pivot, Transpose*)
2. Structural Change of **Subset of Rows and Columns** (*Subtitle*)
3. Change of **Columns** (*Stack, Wide to long*)
4. Change of **Rows and Cells** (*Ffill, Explode*)

Blast Radius from large to small

Task Decomposition

Exploit the sequence of operators

1. Structural Change of **All Rows and Columns** (*Pivot, Transpose*)

Pivot

Name	Alice
Age	17
Name	Bob
Age	19

Name	Age
Alice	17
Bob	19

Transpose

Name	Alice	Bob
Age	17	19

Name	Age
Alice	17
Bob	19

Task Decomposition

Exploit the sequence of operators

2 Structural Change of **Subset of Rows and Columns** (*Subtitle*)

Subtitle

Name	Age
Student	
Alice	17
Bob	19
Teacher	
Claire	32

Name	Age	Role
Alice	17	Student
Bob	19	Student
Claire	32	Teacher

Task Decomposition

Exploit the sequence of operators

3 Change of **Columns** (*Stack, Wide to long*)

Wide_to_long

Name	2013 English	2014 English	2015 English	2013 Math	2014 Math	2015 Math
Alice	A	B	A	B	B	A

Name	Year	Course	Grade
Alice	2013	English	A
Alice	2014	English	B
Alice	2015	English	A
Alice	2013	Math	B
Alice	2014	Math	B
Alice	2015	Math	A

Stack

Name	2013	2014	2015
Alice	A	B	A

Name	Year	Grade
Alice	2013	A
Alice	2014	B
Alice	2015	A

Task Decomposition

Exploit the sequence of operators

4 Change of **Rows and Cells** (*Ffill, Explode*)

Ffill

Role	Name
Student	Alice
	Bob
Teacher	Claire
	David

Role	Name
Student	Alice
Student	Bob
Teacher	Claire
Teacher	David

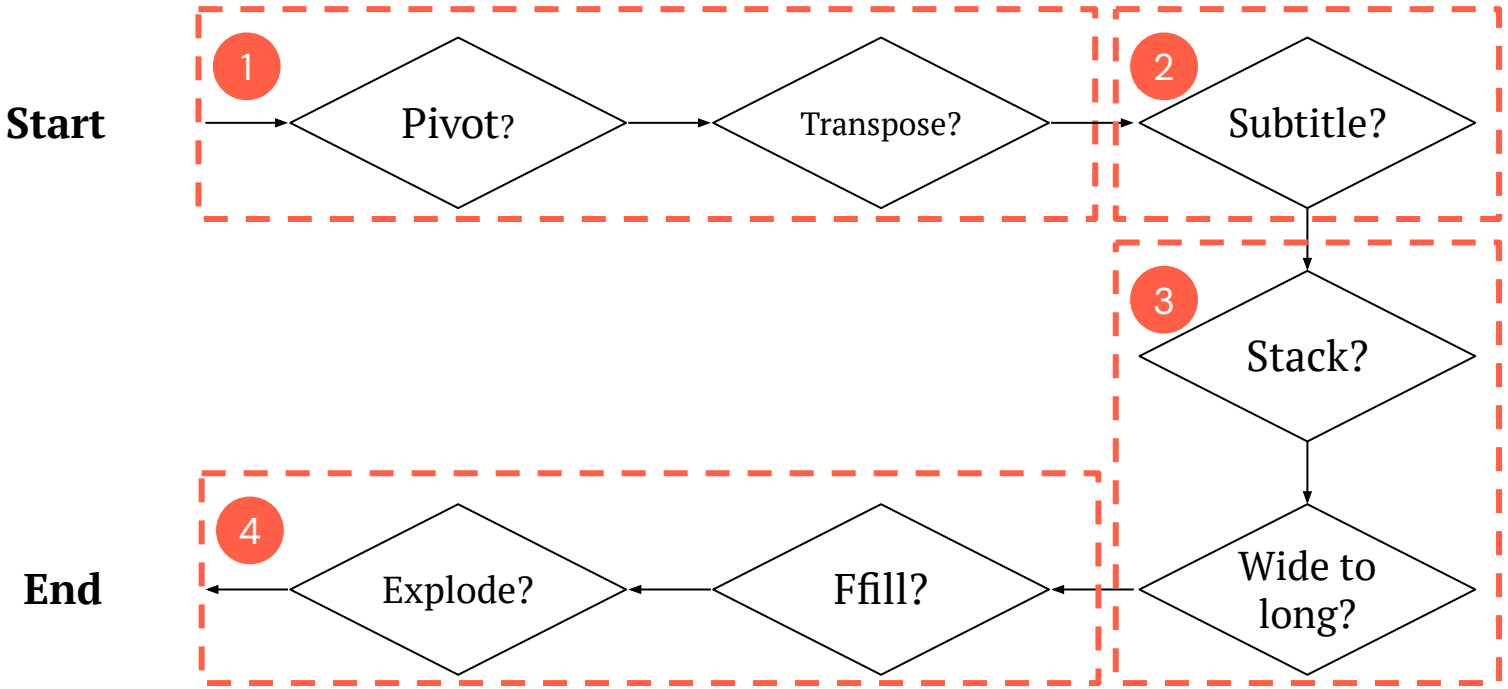
Explode

Teacher	Student
Claire	Alice, Bob
David	Eva, Fiona

Teacher	Student
Claire	Alice
Claire	Bob
David	Eva
David	Fiona

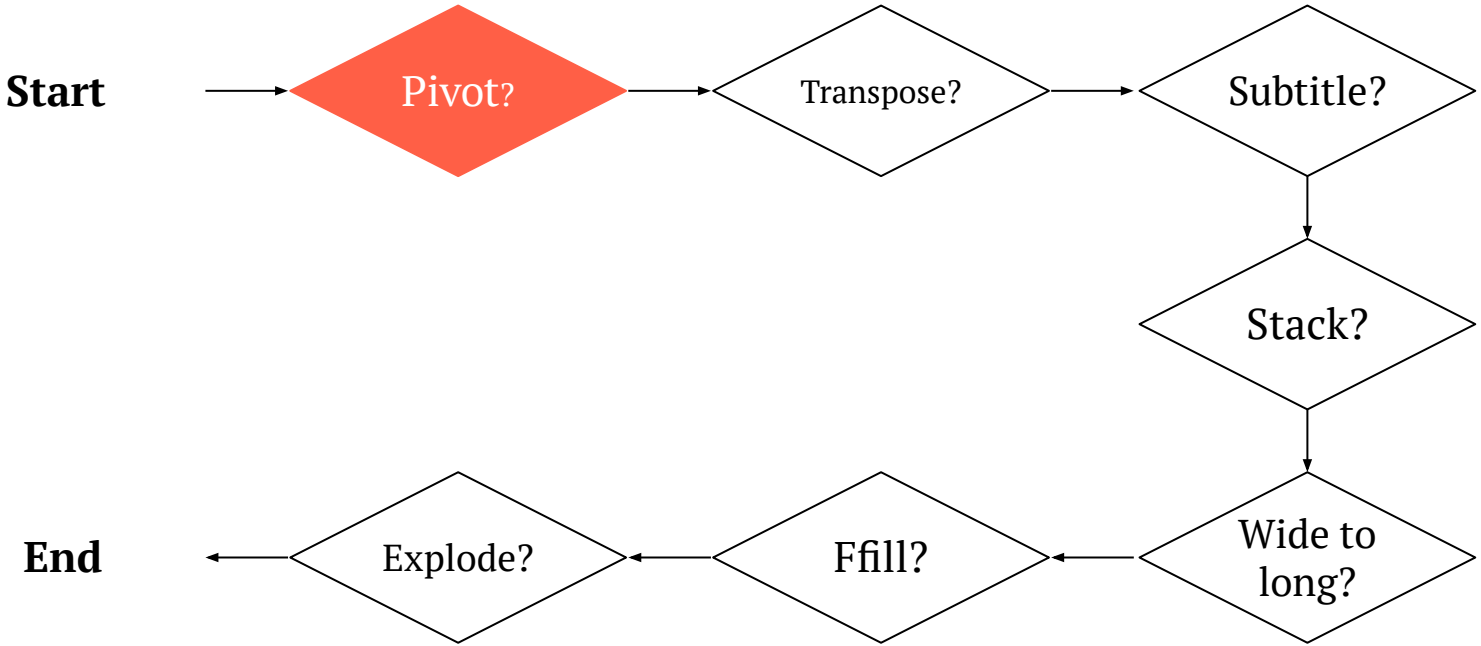
Task Decomposition

Exploit the sequence of operators



Task Decomposition

Exploit the sequence of operators



Task Decomposition

Prompt for Pivot

Example for few-shot learning

Task: Decide if the table needs Pivot
=========
Example Table:
|Name|Alice|
|Age|17|
This table rows follow a pattern:
| Attribute Name | Attribute Value |
"Name" and "Age" should be column headers.
Thus, Pivoting is needed every 2 row.
=========
Input Table:
(input_table_parsed_string)
1. Identify if the pattern exists in the rows.
2. If yes, count the attributes in each group.
=========
Now, return your answer in JSON: {
"reasoning": "The rows mean ...",
"pattern_exist": true/false,
"number_of_attributes_per_group": integer}

CoT

Experiment Results

How much do better prompts/models help?

Method		Acc (%)
Auto-Tables [Li et al.(2023)]		57.0
GPT-3.5	Large Prompt	13.1
	Large Prompt + CoT	
	Decomposed	
	Decomposed + CoT	
GPT-4	Large Prompt	
	Large Prompt + CoT	
	Decomposed	
	Decomposed + CoT	

Experiment Results

How much do better prompts/models help?

Method		Acc (%)
Auto-Tables [Li et al.(2023)]		57.0
GPT-3.5	Large Prompt	13.1
	Large Prompt + CoT	8.2
	Decomposed	0
	Decomposed + CoT	3.2
GPT-4	Large Prompt	
	Large Prompt + CoT	
	Decomposed	
	Decomposed + CoT	

Prompt engineering degrades GPT-3.5

Experiment Results

How much do better prompts/models help?

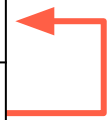
Method		Acc (%)
Auto-Tables [Li et al.(2023)]		57.0
GPT-3.5	Large Prompt	13.1
	Large Prompt + CoT	8.2
	Decomposed	0
	Decomposed + CoT	3.2
GPT-4	Large Prompt	46.3
	Large Prompt + CoT	
	Decomposed	
	Decomposed + CoT	

GPT-4 greatly improves over GPT-3.5

Experiment Results

How much do better prompts/models help?

Method		Acc (%)
Auto-Tables [Li et al.(2023)]		57.0
GPT-3.5	Large Prompt	13.1
	Large Prompt + CoT	8.2
	Decomposed	0
	Decomposed + CoT	3.2
GPT-4	Large Prompt	46.3
	Large Prompt + CoT	39.3
	Decomposed	
	Decomposed + CoT	



CoT still degrades GPT-4 over a large prompt

Experiment Results

How much do better prompts/models help?

Method		Acc (%)
Auto-Tables [Li et al.(2023)]		57.0
GPT-3.5	Large Prompt	13.1
	Large Prompt + CoT	8.2
	Decomposed	0
	Decomposed + CoT	3.2
GPT-4	Large Prompt	46.3
	Large Prompt + CoT	39.3
	Decomposed	60.1
	Decomposed + CoT	

Task Decomposition greatly improve GPT-4

Experiment Results

How much do better prompts/models help?

Method		Acc (%)
Auto-Tables [Li et al.(2023)]		57.0
GPT-3.5	Large Prompt	13.1
	Large Prompt + CoT	8.2
	Decomposed	0
	Decomposed + CoT	3.2
GPT-4	Large Prompt	46.3
	Large Prompt + CoT	39.3
	Decomposed	60.1
	Decomposed + CoT	74.6

CoT improves Task Decomposition for GPT-4

Experiment Results

How much do better prompts/models help?

Method		Acc (%)
Auto-Tables [Li et al.(2023)]		57.0
GPT-3.5	Large Prompt	13.1
	Large Prompt + CoT	8.2
	Decomposed	0
	Decomposed + CoT	3.2
GPT-4	Large Prompt	46.3
	Large Prompt + CoT	39.3
	Decomposed	60.1
	Decomposed + CoT	74.6

25.4% errors?

Error Analysis

Out of the 25.4% errors...

Category		%
Ambiguity	Transpose	1.2
	Stack	7.4
	Explode	3.3
Content Filtering		0.4
True Mistake		13.1

~ half due to ambiguity

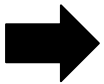
Error Analysis

Out of the 25.4% errors, ~ half due to ambiguity

1. Transpose Ambiguity (1.2%)

Name	Alice	Bob
Age	17	19

Transpose



Name	Age
Alice	17
Bob	19

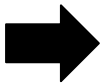
Error Analysis

Out of the 25.4% errors, ~ half due to ambiguity

1. Transpose Ambiguity (1.2%)

	Asian	Black	Hispanic
Brown	14%	6%	10%
Columbia	15%	8%	13%
Cornell	17%	6%	10%

Transpose



	Brown	Columbia	Cornell
Asian	14%	15%	17%
Black	6%	8%	6%
Hispanic	10%	13%	10%

Table shows a **Pivot View**

- **Ground Truth:** Transpose
- **GPT-4:** Not Transpose

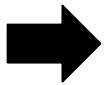
Error Analysis

Out of the 25.4% errors, ~ half due to ambiguity

2. Stack Ambiguity (7.4%)

Name	2013	2014	2015
Alice	A	B	A

Stack



Name	Year	Grade
Alice	2013	A
Alice	2014	B
Alice	2015	A

Error Analysis

Out of the 25.4% errors, ~ half due to ambiguity

2. Stack Ambiguity (7.4%)

Team	Win	Loss
Boston Celtics	17	4
LA Lakers	16	15

Stack

Team	Result	Value
Boston Celtics	Win	17
Boston Celtics	Loss	4
LA Lakers	Win	16
LA Lakers	Loss	15

- **Ground Truth:** Not Stack
- **GPT-4:** Stack

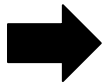
Error Analysis

Out of the 25.4% errors, ~ half due to ambiguity

3. Explode Ambiguity (3.3%)

Teacher	Student
Claire	Alice, Bob
David	Eva, Fiona

Explode



Teacher	Student
Claire	Alice
Claire	Bob
David	Eva
David	Fiona

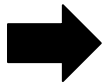
Error Analysis

Out of the 25.4% errors, ~ half due to ambiguity

3. Explode Ambiguity (3.3%)

Name	Population
Battery Park City, Lower Manhattan	20088

Explode



Name	Population
Battery Park City	20088
Lower Manhattan	20088

- **Ground Truth:** Explode
- **GPT-4:** Not Explode

Error Analysis

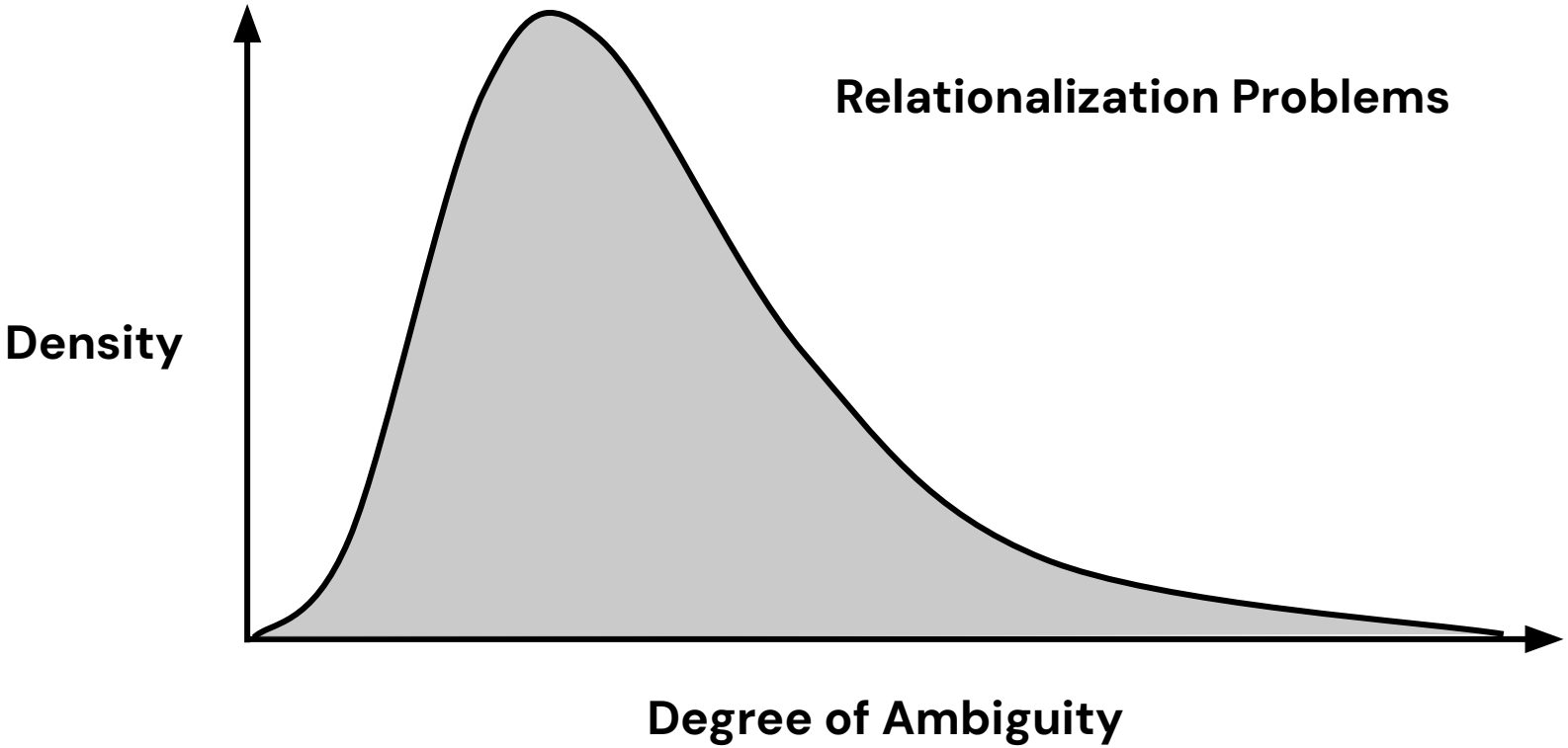
Out of the 25.4% errors, ~ half due to ambiguity

4. Content Filtering (0.4%)

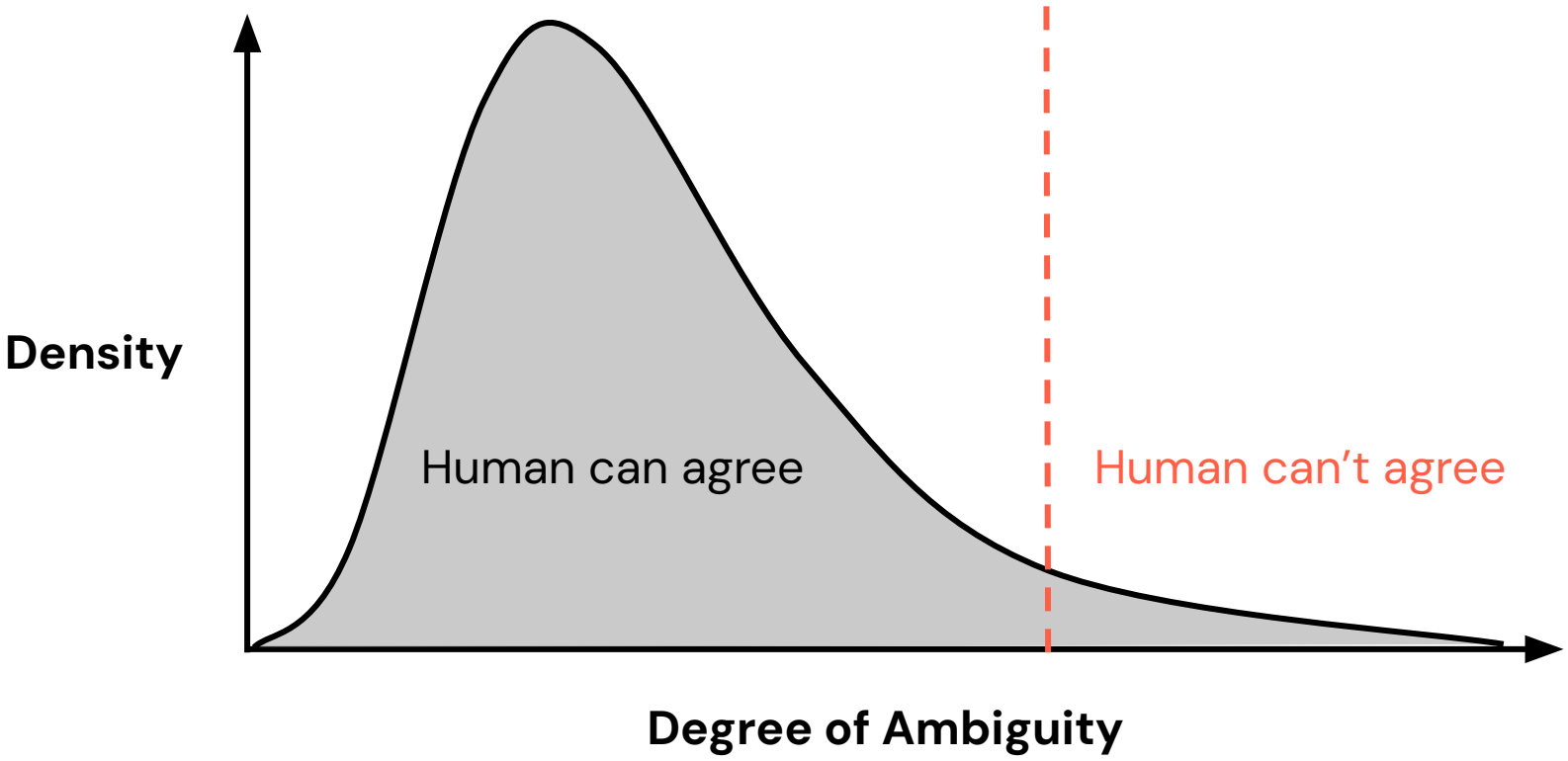
Band - Title	Label	Price
Animals Killing People - Human Hunting Season	1	0
Avulsed - Gorespattered Suicide	0	0

The response was filtered due to the prompt triggering OpenAI's content management policy. Please modify your prompt and retry.

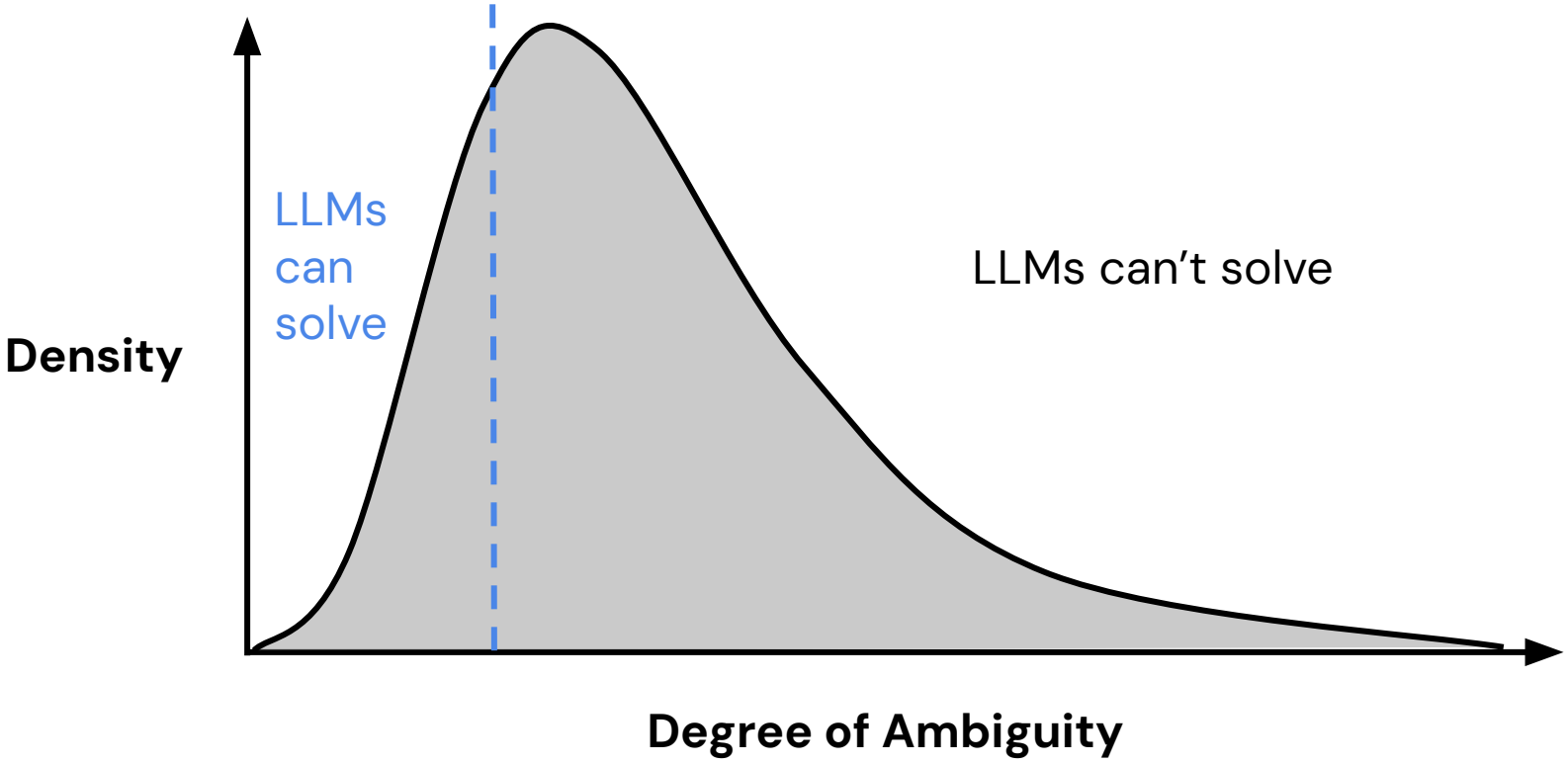
Parting Thoughts



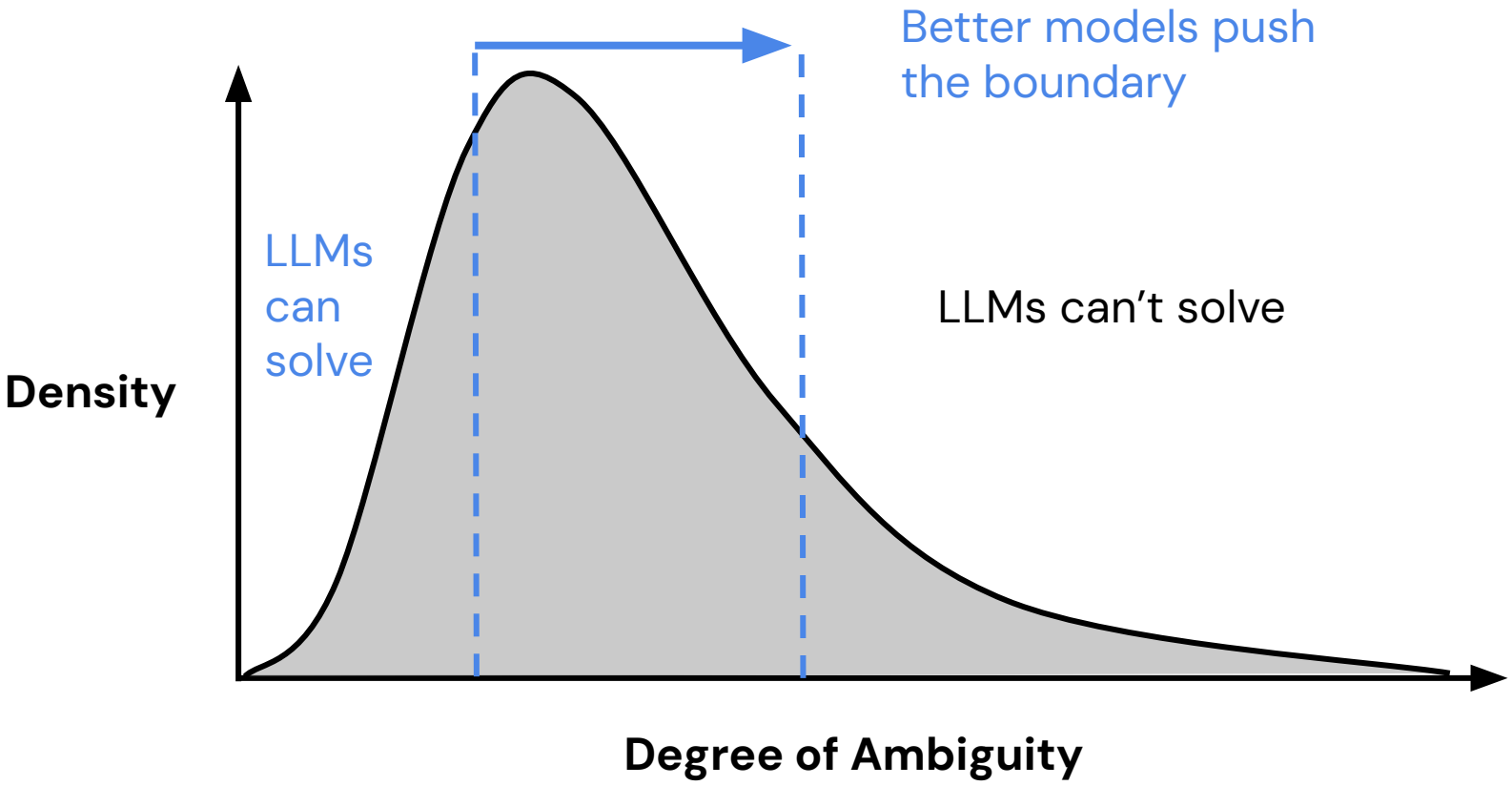
Parting Thoughts



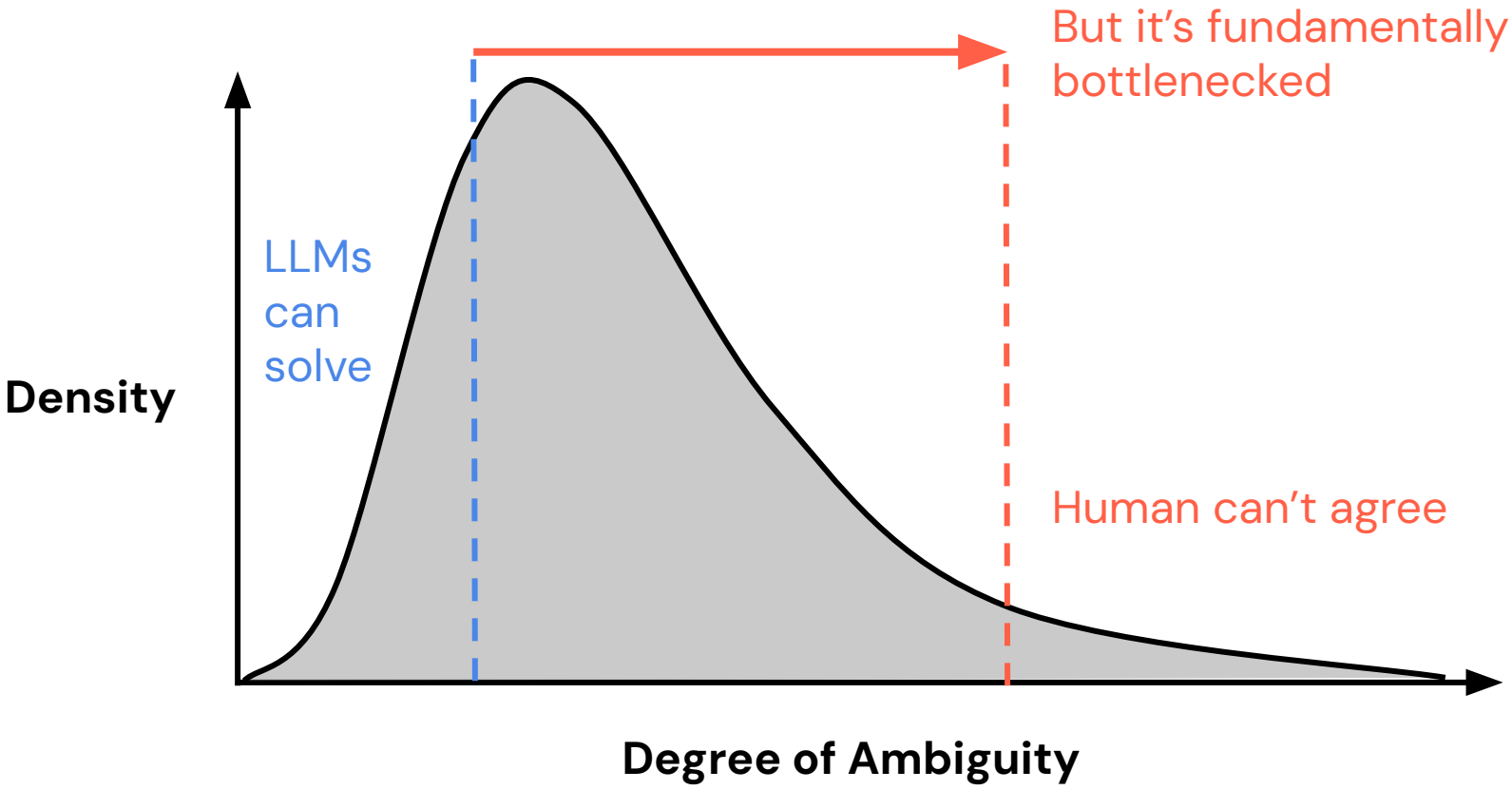
Parting Thoughts



Parting Thoughts



Parting Thoughts



Parting Thoughts