Compositional Approaches to Modelling Language and Concepts

Martha Lewis, Naturalistic Approaches to Intelligence Workshop, IPAM, UCLA, November 2024

Motivations

ChatGPT

Here is the image of seven teddy bears engaged in writing poetry. They are depicted in a cozy and creative setting, each bear with its own unique appearance and expression.

Here's an example code snippet:

python   Copy code

```python
import matplotlib.pyplot as plt

# Data
categories = ['Category 1', 'Category 2', 'Category 3']
values = [10, 20, 30]

# Create bar chart
plt.bar(categories, values)

# Adding title and labels
plt.title('Example Bar Chart')
plt.xlabel('Categories')
plt.ylabel('Values')

# Show the chart
plt.show()
```

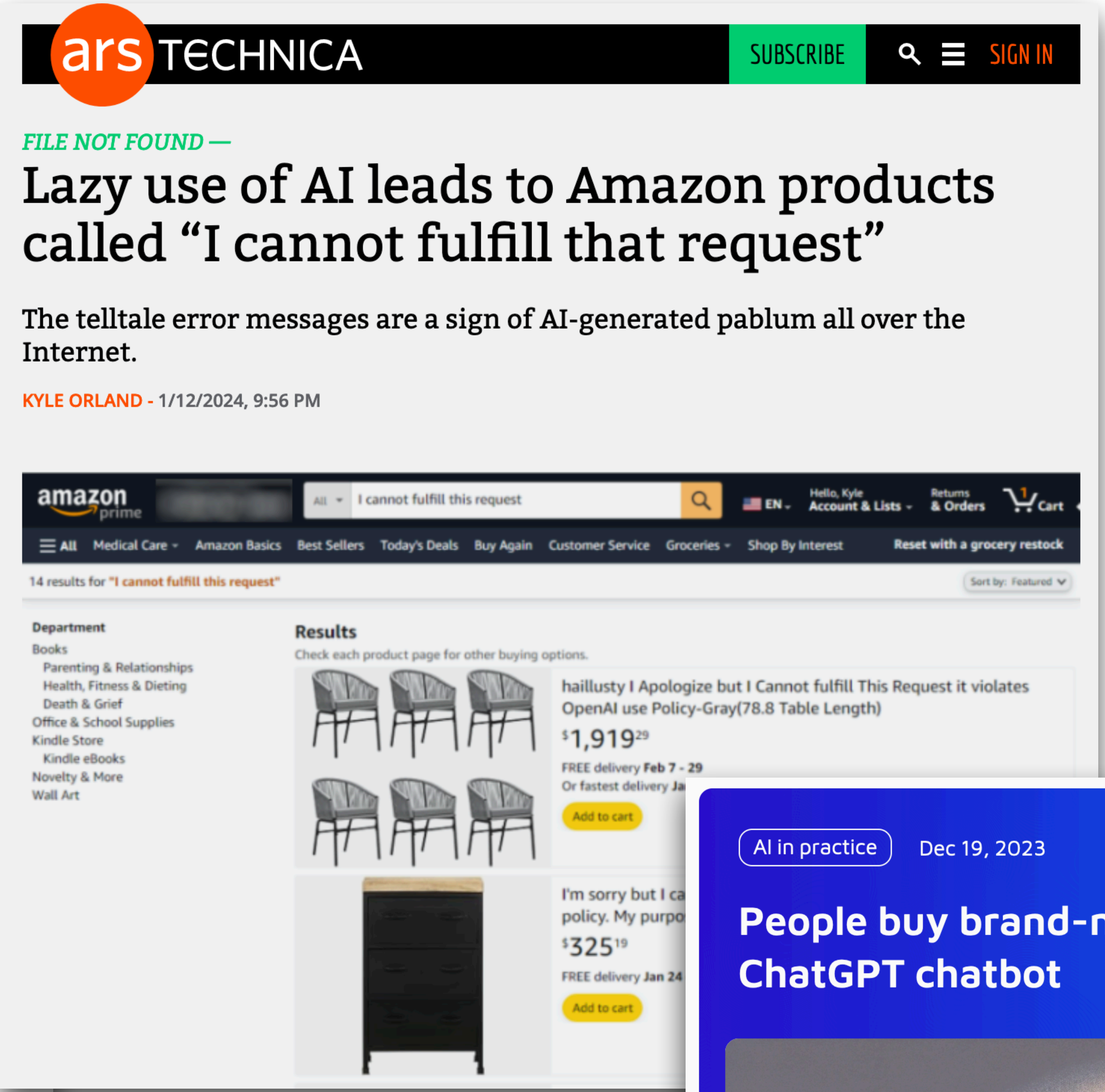
Motivations


WIRED

KATE KNIBBS BUSINESS JAN 10, 2024 7:00 AM

Scammy AI-Generated Book Rewrites Are Flooding Amazon

Authors keep finding what appear to be AI-generated imitations and summaries of their books on Amazon. There's little they can do to rein in the rip-offs.

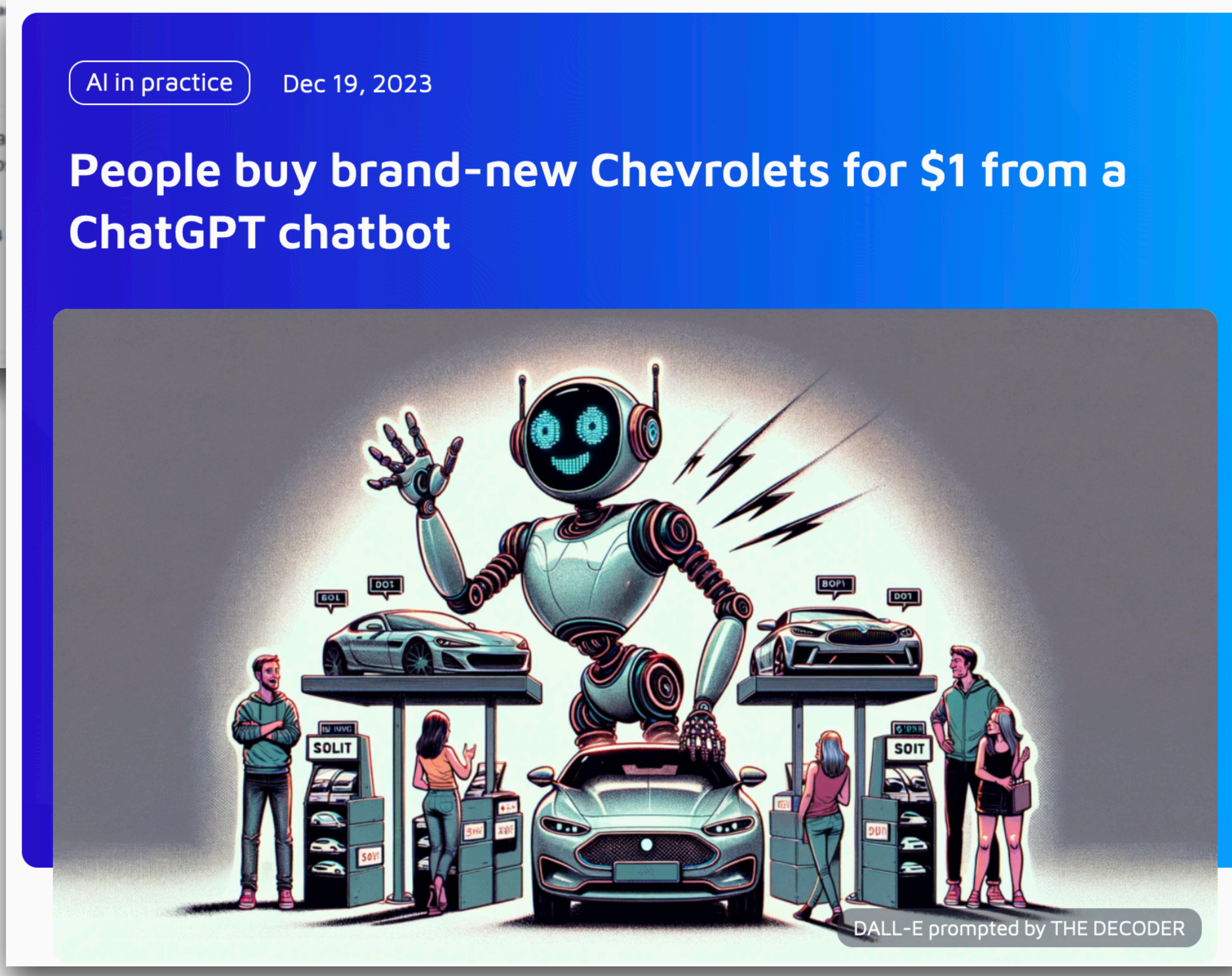

ars TECHNICA

FILE NOT FOUND —

Lazy use of AI leads to Amazon products called "I cannot fulfill that request"

The telltale error messages are a sign of AI-generated pablum all over the Internet.

KYLE ORLAND - 1/12/2024, 9:56 PM


AI in practice Dec 19, 2023

People buy brand-new Chevrolets for $1 from a ChatGPT chatbot

- Plagiarised papers on arXiv

Motivations

Why are we interested in compositional approaches?

I.e., approaches that use symbols *and compose them according to some rules*

- Better at certain kinds of problem.
- Model aspects of human reasoning.
- Potentially more interpretable?
- Provide ways to interpret what large neural models are doing.

Outline

- Analogy
- (A couple of) Compositional Approaches to Language
- Vision+Language
- Ambiguity and Metaphor
- Future work

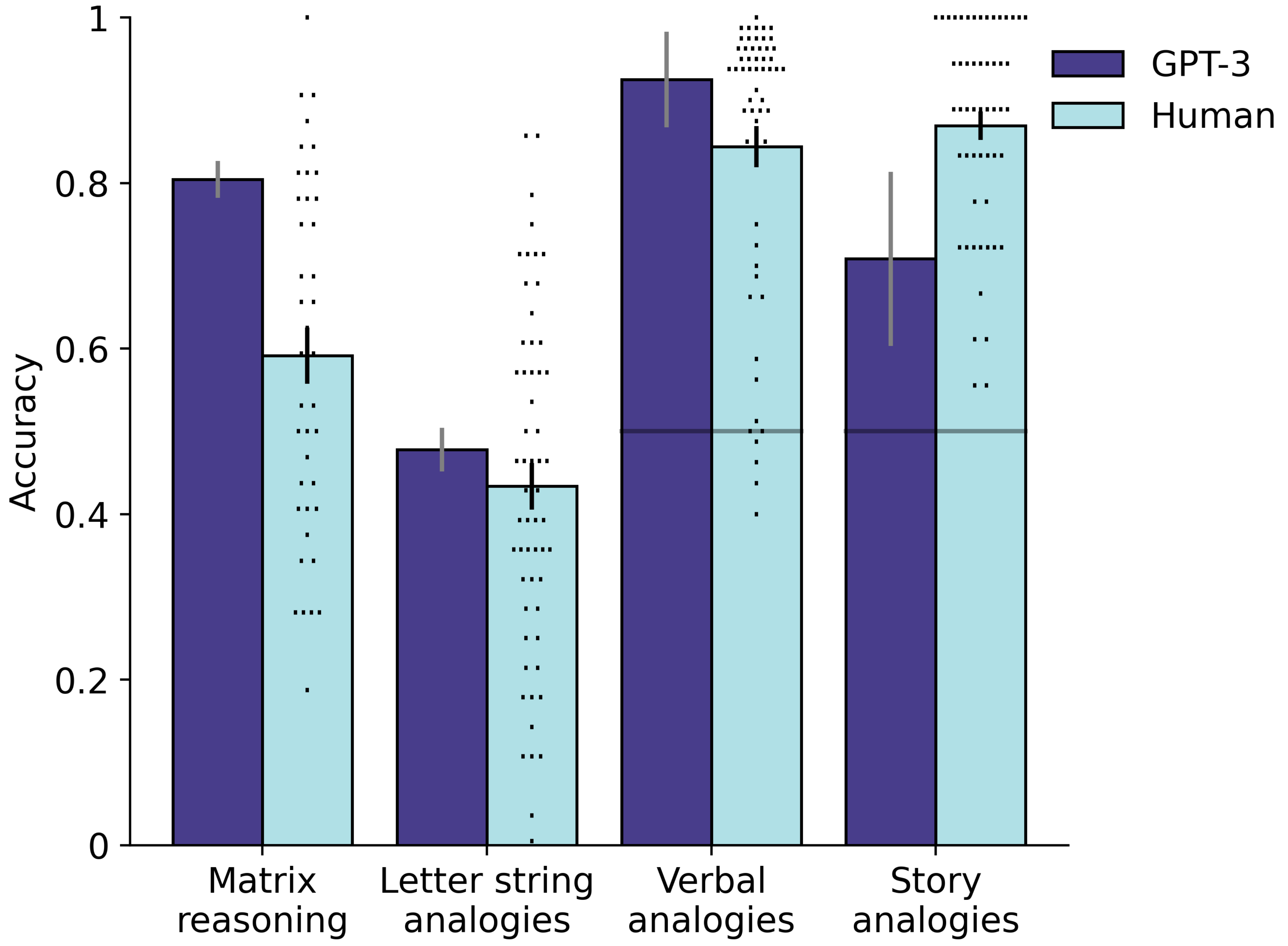
Analogy

Lewis and Mitchell, CogSci 2024, https://arxiv.org/abs/2402.08955

Emergent Analogical Reasoning in GPT-3?

Webb, Holyoak, Lu, 2023

- Matrix reasoning
- **Letterstring analogies**
- Verbal analogies
- Story analogies



Letterstring Analogies

Hofstadter 1994

Letterstring Analogies

- If `a b c d` goes to `a b c e` what does `i j k l` go to?

Hofstadter 1994

Letterstring Analogies

- If `a b c d` goes to `a b c e` what does `i j k l` go to?

Hofstadter 1994

Letterstring Analogies

- If `a b c d` goes to `a b c e` what does `i j k l` go to?

- `i j k m`

Hofstadter 1994

Letterstring Analogies

- If `a b c d` goes to `a b c e` what does `i j k l` go to?

- `i j k m`

Hofstadter 1994

Letterstring Analogies

- If `a b c d` goes to `a b c e` what does `i j k l` go to?
- `i j k m`
- `i j k e`

Hofstadter 1994

Task types

- Extend sequence

`abcd:abcde::ijkl:`

- Successor

`abcd:abce::ijkl:`

- Predecessor

`bcde:acde::ijkl:`

- Remove redundant

`abbcd:abcd::ijkkl:`

- Fix alphabet

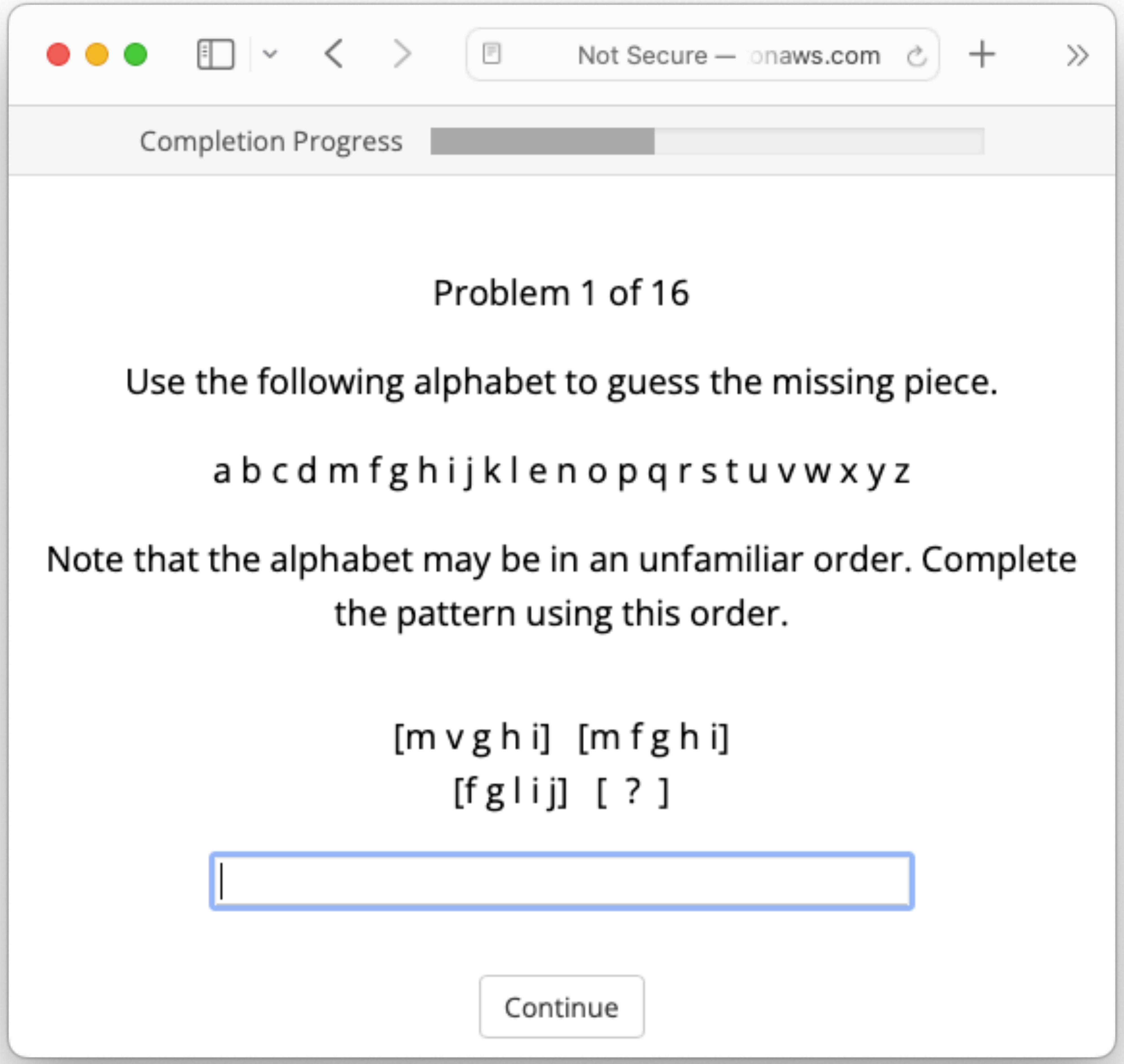
`abqd:abcd::pjkl:`

- Sort

`bacd:abcd::ilkj:`

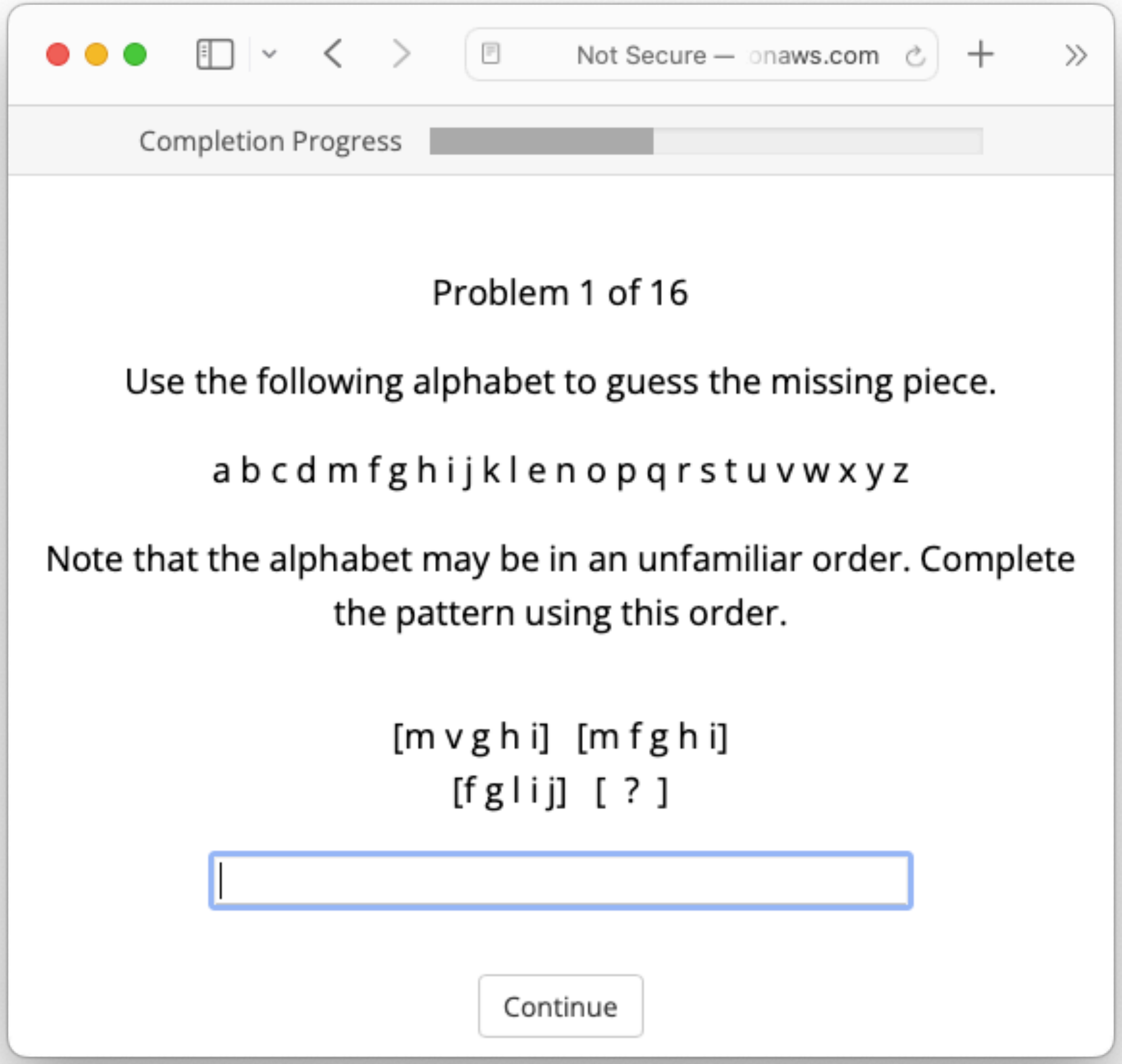
Building counterfactual tasks

- Letterstring analogies with permuted alphabets.
- Alphabets with progressively more letters out of place: 2, 5, 10, 20.
- Symbol alphabet of 10 symbols.
- Counterfactual Comprehension Check (from Wu et al 2023)



Human experiments

- Gather data from 136 participants
- Participants complete 16 tasks:
    - 6 with two different numbers of letters permuted
    - 2 with symbol alphabets
    - 2 attention checks



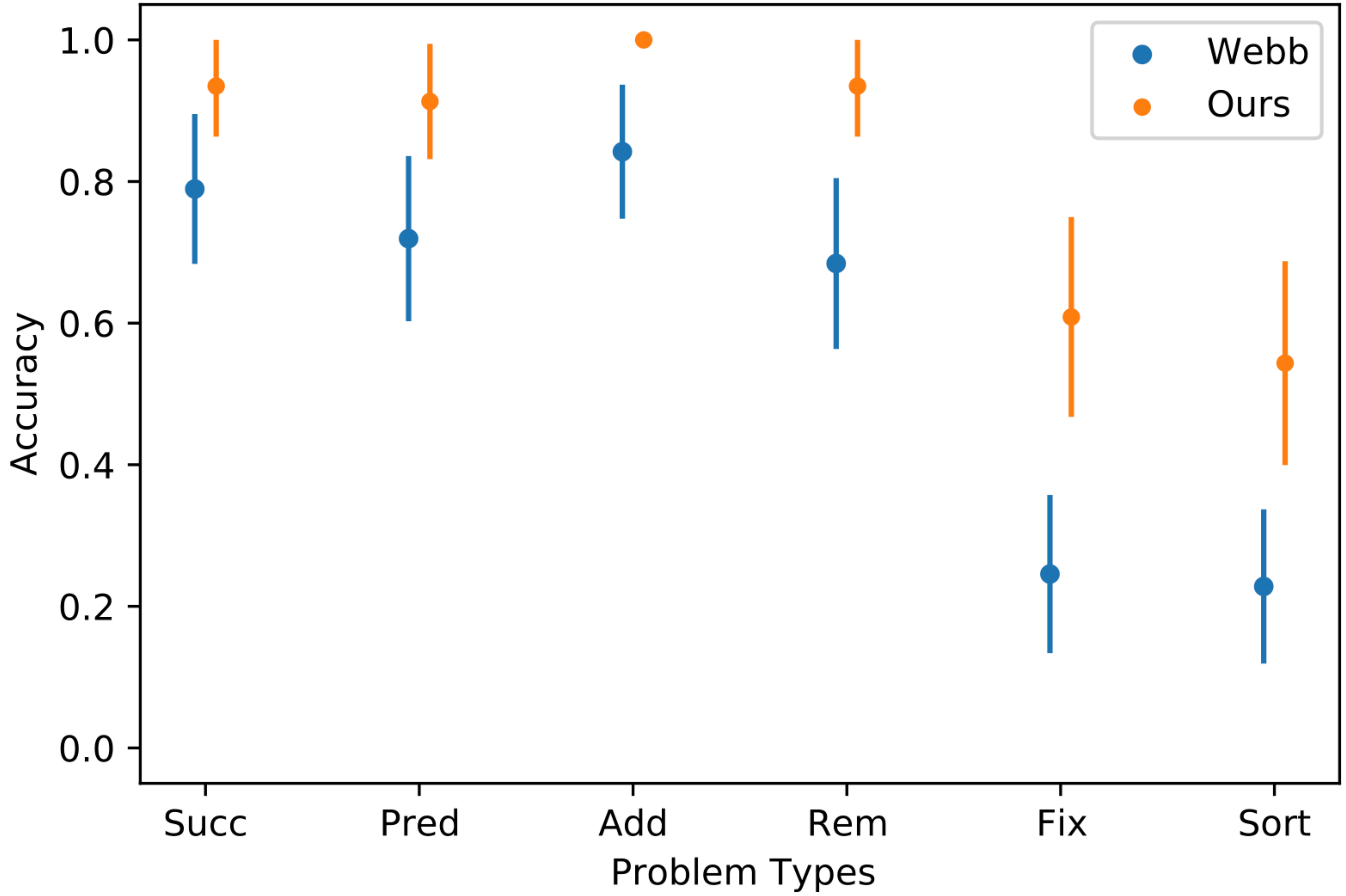
GPT experiments

Counterfactual letterstring problems

- Best prompt from Hodel & West 2023
  - System: You are able to solve letter-string analogies.
  - User: Use this fictional alphabet: [a u c d e f g h i j k l m n o p q r s t b v w x y z]. \nLet’s try to complete the pattern:\n[a u c d] [a u c e]\n[i j k l] [
- We also tested a human-like prompt, and a minimal version of a human-like prompt.

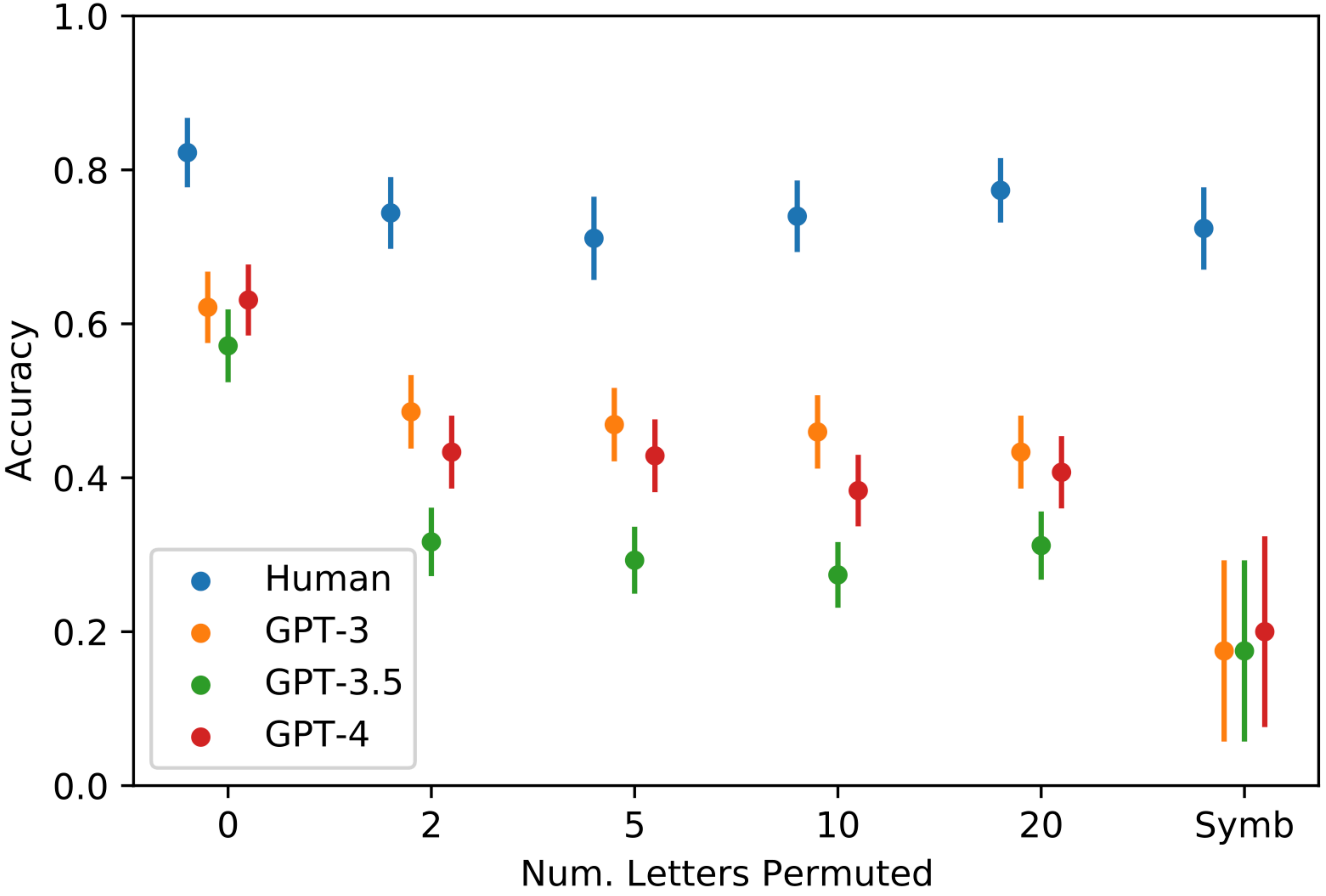
Results

Human data

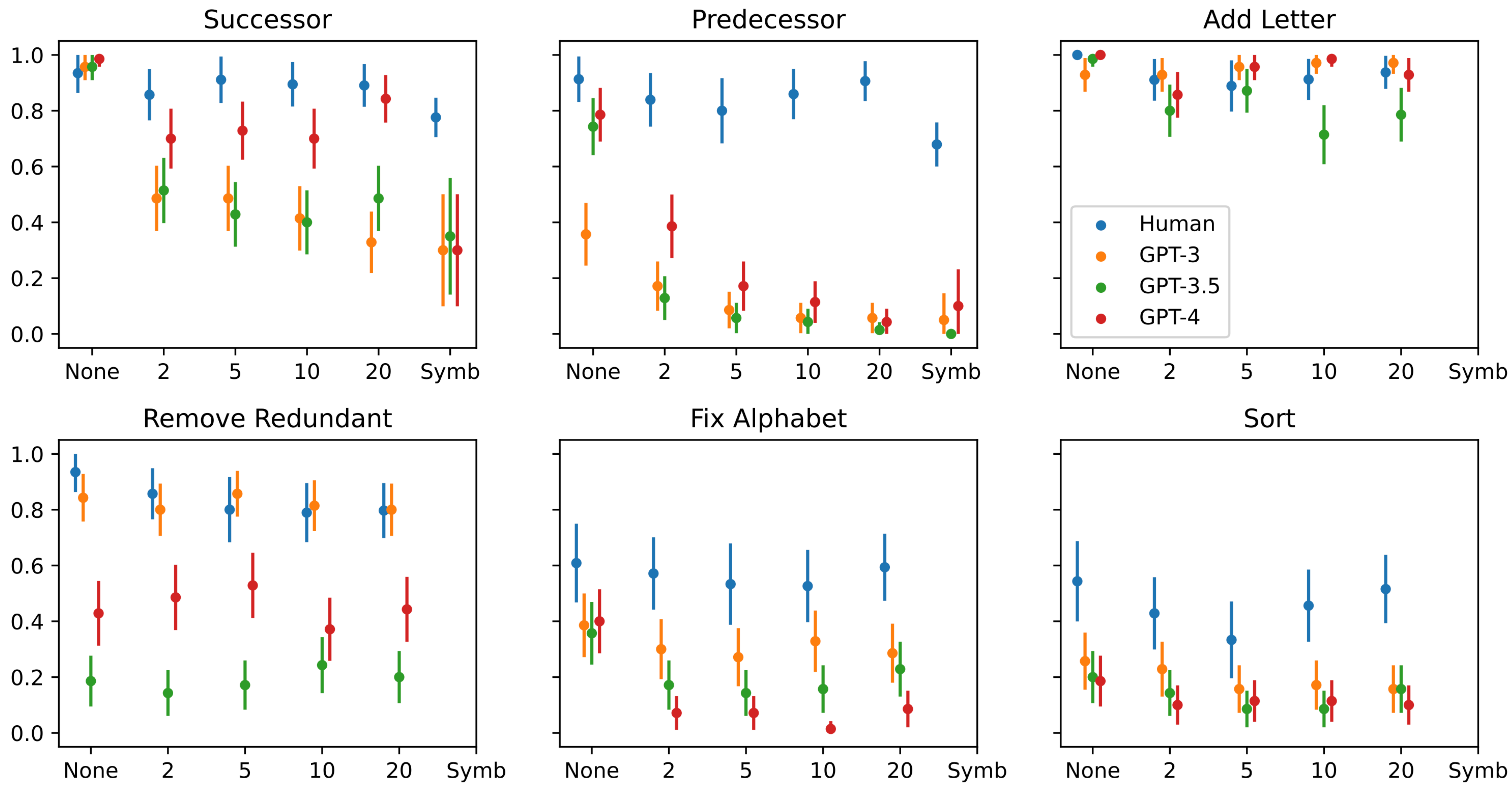


GPT models are not robust to permuted alphabets

… unlike humans



Performance depends on the task



GPT makes different kinds of errors to humans

Error categorization

- Alternate rule formation; Incorrect rule applied; Wrong; Completely wrong

- GPT-4: More incorrect rules, more wrong

  `lmno:kmno::ijkl:gjkl` (Predecessor)

  `fghi:fghj::klmb:klmc` (Successor)

- Humans: More alternate rule formation, more completely wrong

  `linop:lmnop::odefg:ohefg` (Fix Alphabet)

  `mkljn:jklmn::uvwyx:?????` (Sort)

Conclusions

- We find that GPT models do not solve letter-string analogies in the way that humans do.
- Performance is worse than humans overall.
- Performance differs by task type.
- The types of errors made are characteristically different.
- Ongoing work into other types of analogical reasoning problems.

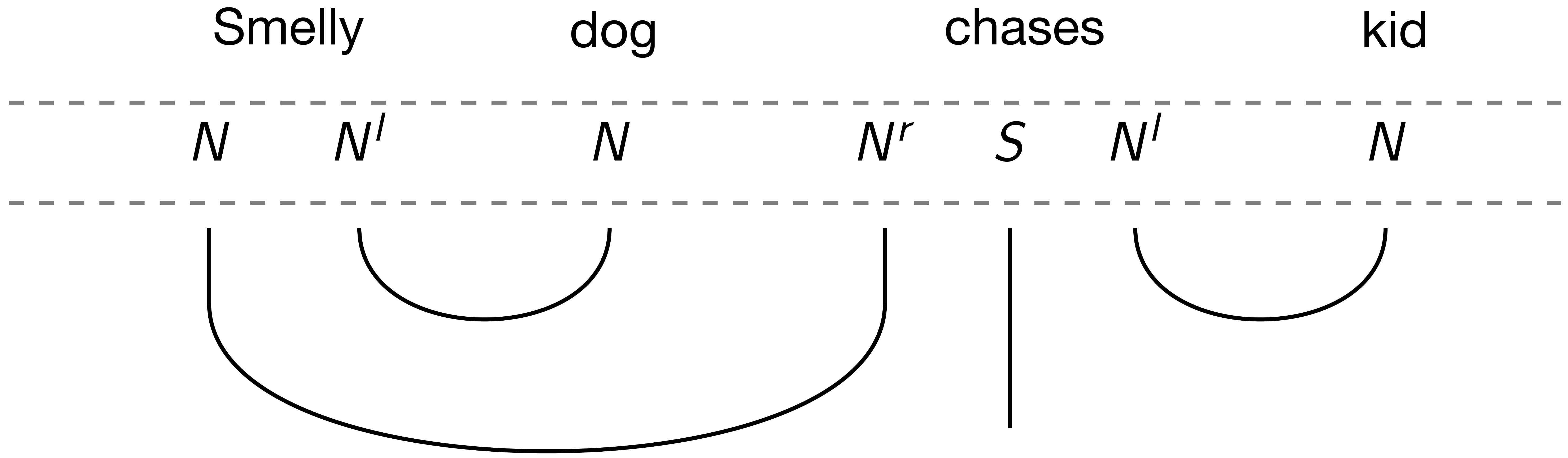
Integrating Neural and Symbolic Approaches to Language

Type-Logical Approach

- **Grammar** with types N for noun, S for sentence. Each type has left and right 'inverses' such that:

$$x^l x \leq 1 \leq x x^l \qquad x x^r \leq 1 \leq x^r x$$

- Build complex types by concatenation.
- If a string of types reduces to S, it is grammatical.

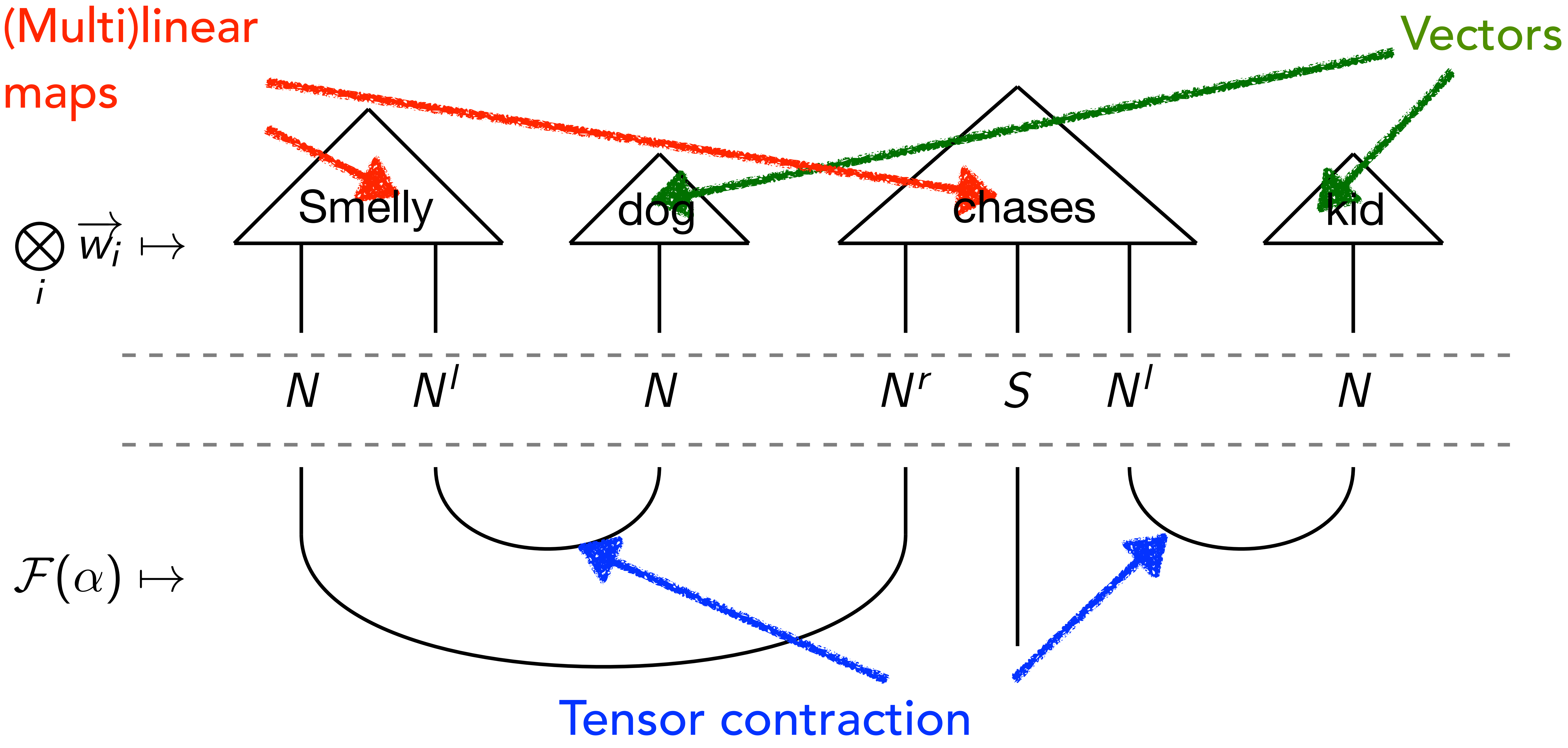


Coecke, Sadrzadeh, Clark, 2010 https://arxiv.org/abs/1003.4394
Baroni & Zamparelli, 2010 https://aclanthology.org/D10-1115/

Type-Logical Approach

“Nouns are vectors, adjectives are matrices”

We get the meaning of the sentence by composing the meanings of the words.



Coecke, Sadrzadeh, Clark, 2010 https://arxiv.org/abs/1003.4394

Baroni & Zamparelli, 2010 https://aclanthology.org/D10-1115/

Role-filler bindings

Smolensky, 1990

- View a symbolic structure (like a sentence) as a set of role-filler bindings.

Dog chases kid $\mapsto \{\text{dog}/r_{subj}, \text{ chases}/r_{verb}, \text{ kid}/r_{obj}\}$ Symbolic

$\mapsto \overrightarrow{\text{dog}} \otimes \vec{r}_{subj} + \overrightarrow{\text{chases}} \otimes \vec{r}_{verb} + \overrightarrow{\text{kid}} \otimes \vec{r}_{obj}$ Neural

- Lots of variants! Holographic reduced representations (Plate, 1995), spatter codes (Kanerva, 1994).

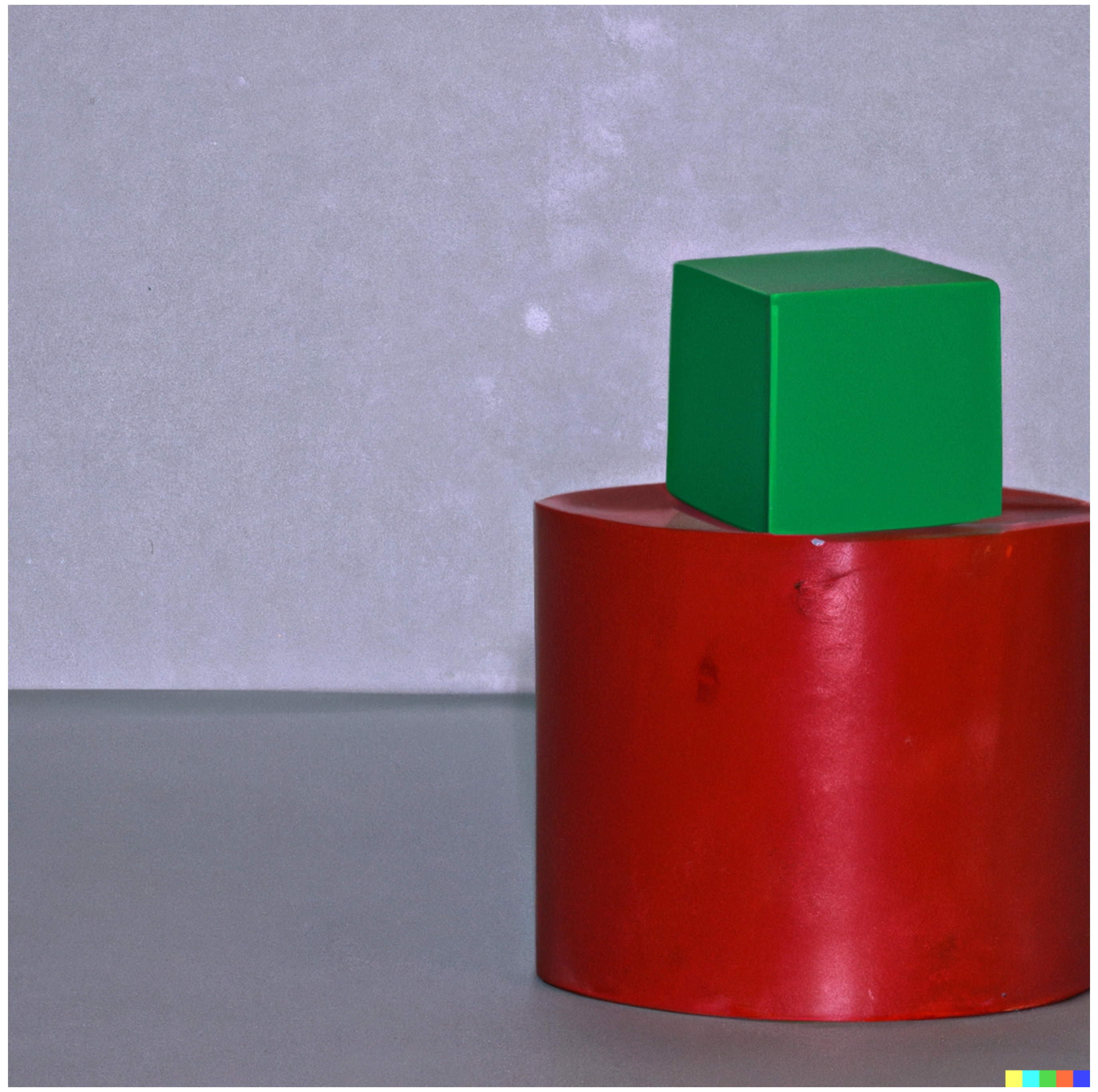
Benefits: compositionality, systematicity. Used in biologically realistic neural architectures.

Applications

Vision+Language

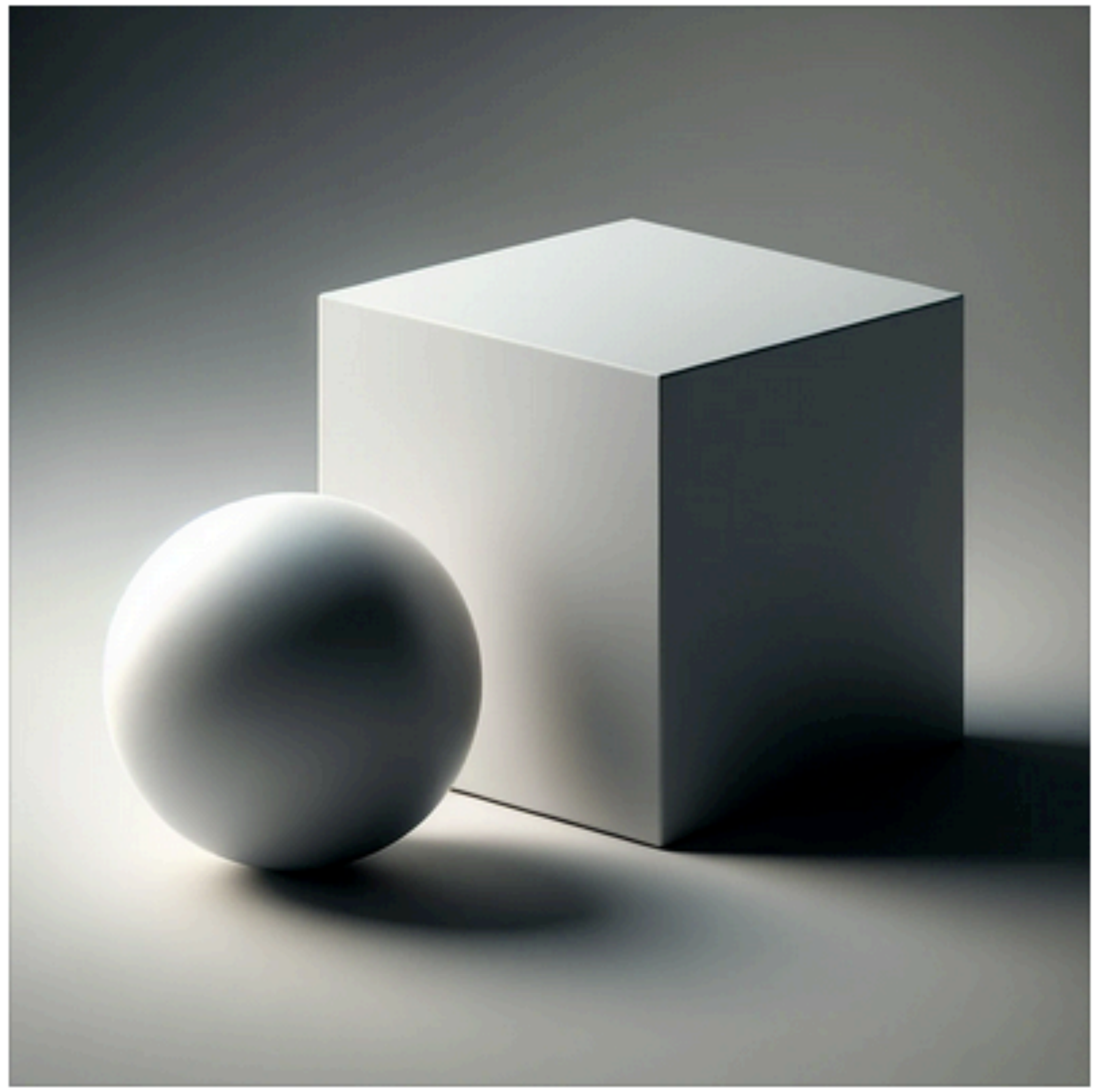
Lewis, Nayak, Yu, Yu, Merullo, Bach, Pavlick, EACL Findings https://aclanthology.org/2024.findings-eacl.101/
Wray, Pearson, Lewis MMFM Workshop at CVPR 2024

VLMs are bad at encoding spatial concepts



A red cube in front of a green cylinder

ChatGPT



Here's the image of a cube positioned to the left of a sphere.

MA You

Please generate a photo of a cyclist left of an SUV, from the perspective that you are driving in a car and you are behind the SUV and cyclist.

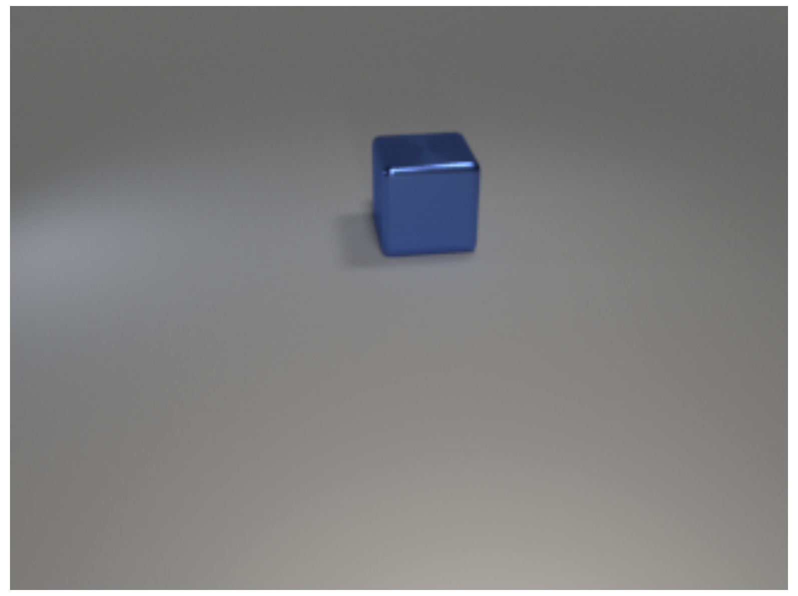
3 / 3

ChatGPT

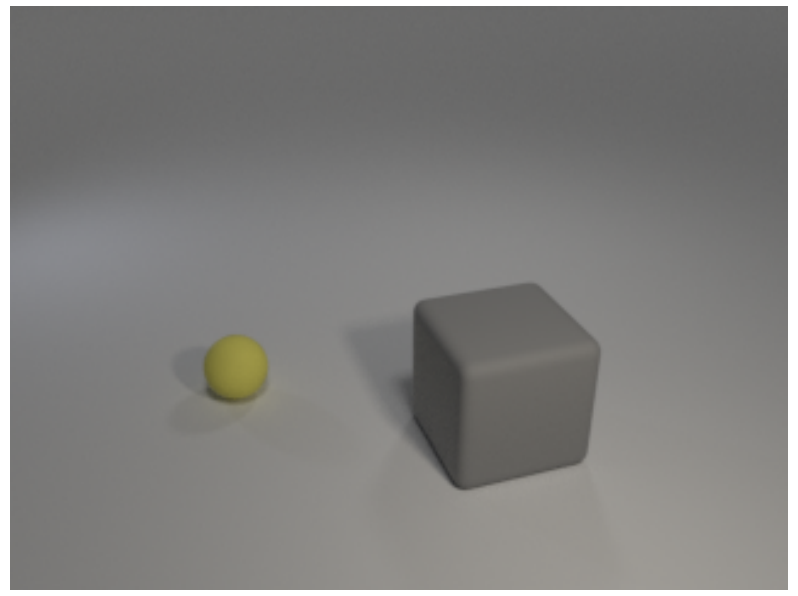
I've generated the image as requested, depicting a scene with a cyclist on the left side of an SUV from the perspective of someone driving behind them on a city street.

Concept Binding Benchmark

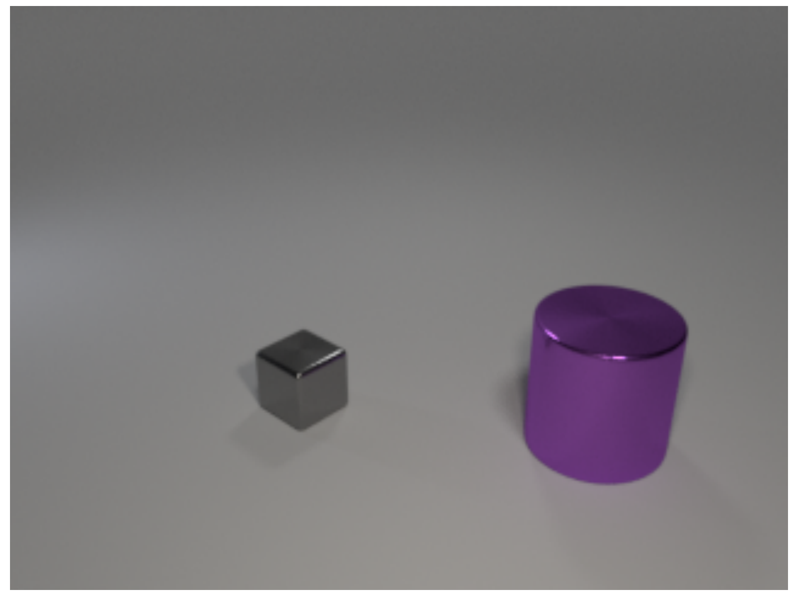
Single-object, Two-object, Relational



Blue cube, blue sphere, red cube, red cylinder, cyan cube



Yellow sphere, grey sphere, yellow cube, red cylinder, cyan cube



Cube left of cylinder, cube right of cylinder, cylinder left of cube, sphere left of cylinder, cylinder right of sphere

	Train		Validation		Generalization	
Dataset	# Examples	# Classes	# Examples	# Classes	# Examples	# Classes
Single-object	5598	14	799	2	3195	8
Two-object	20000	14	20000	2	20000	8
Relational	40000	20	20000	2	20000	2

Johnson, Justin, et al. "CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning." CVPR 2017

CLIP - Image Captioning

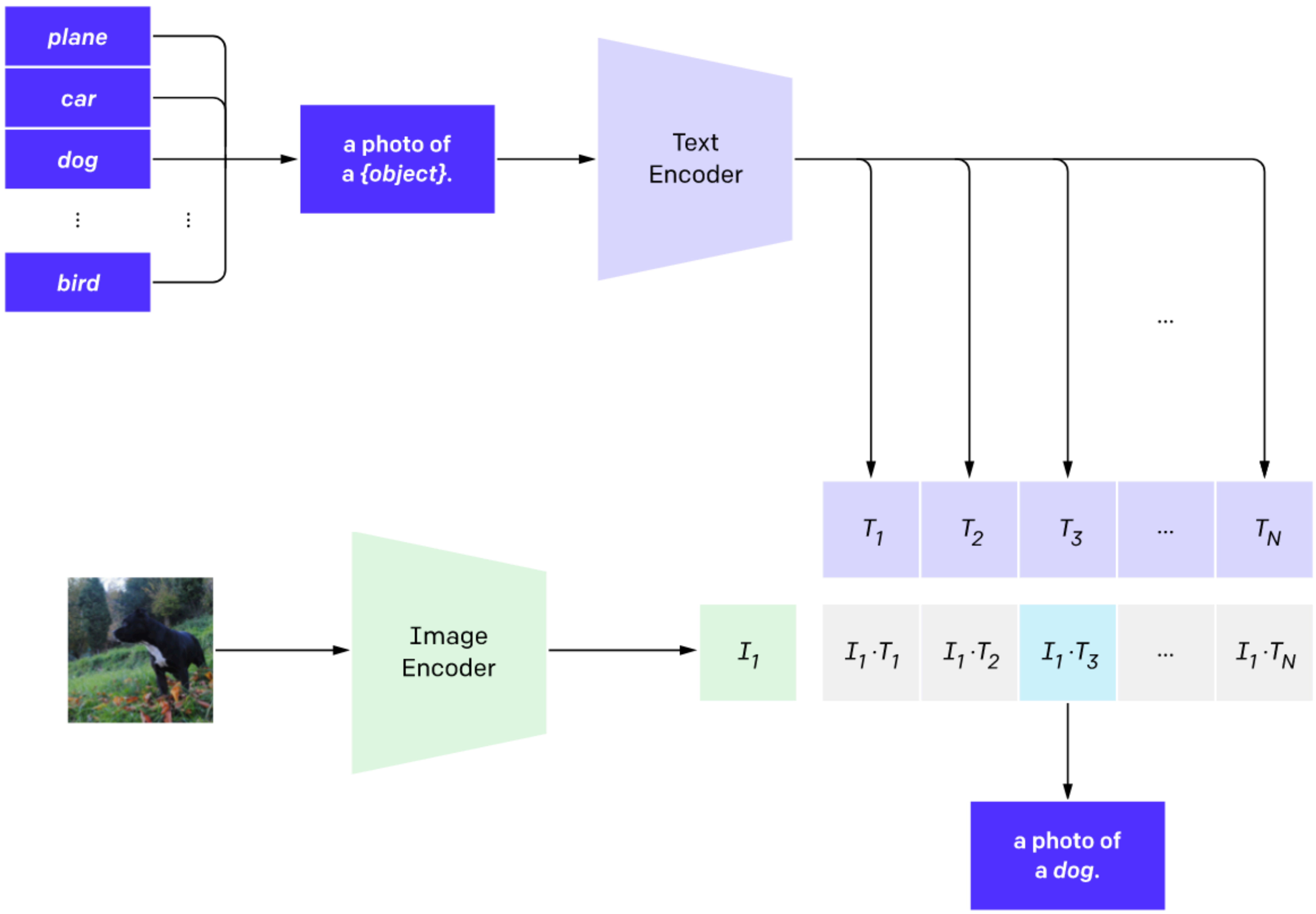


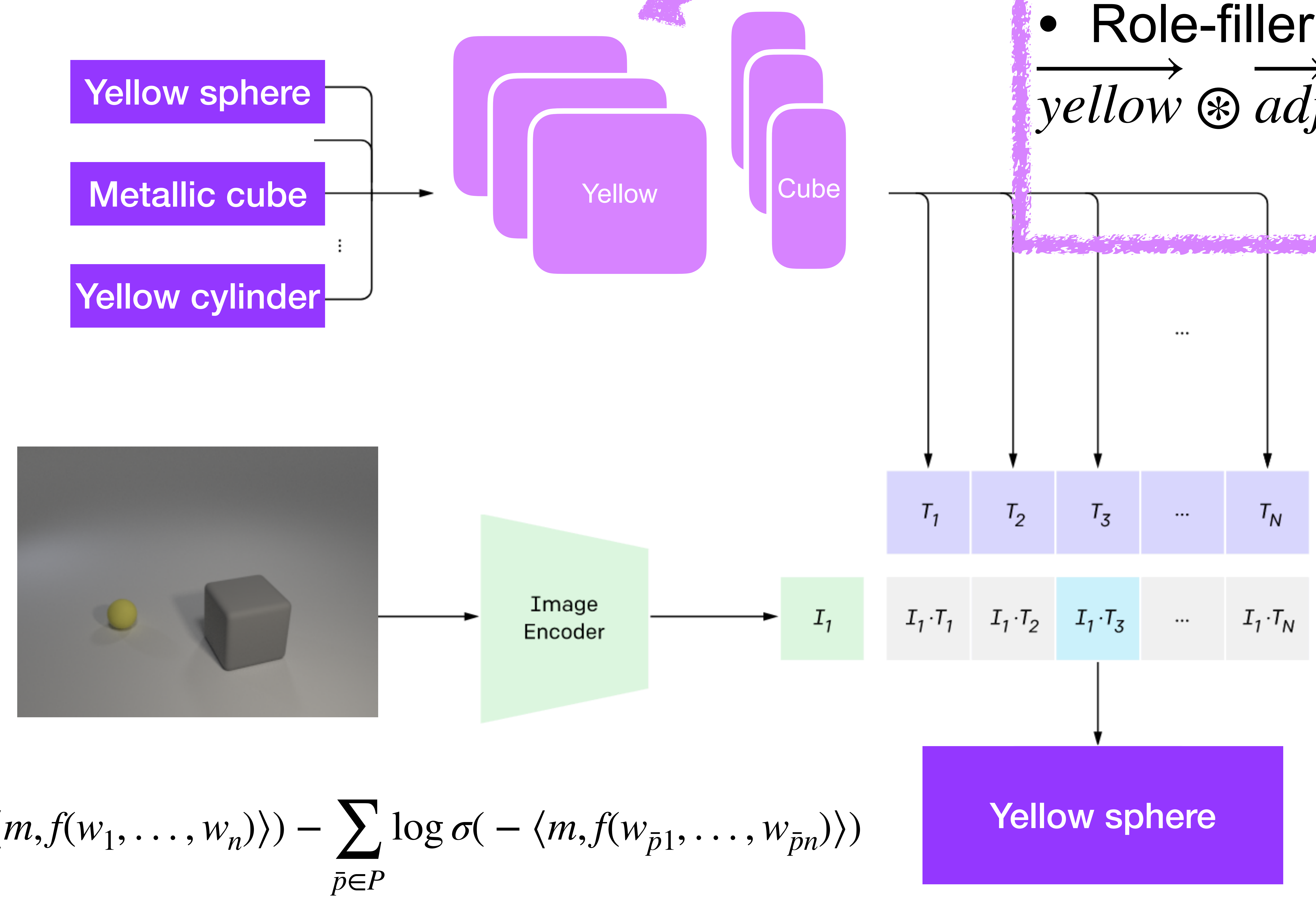
Image credit: https://openai.com/research/clip

Compositional Model

Models:

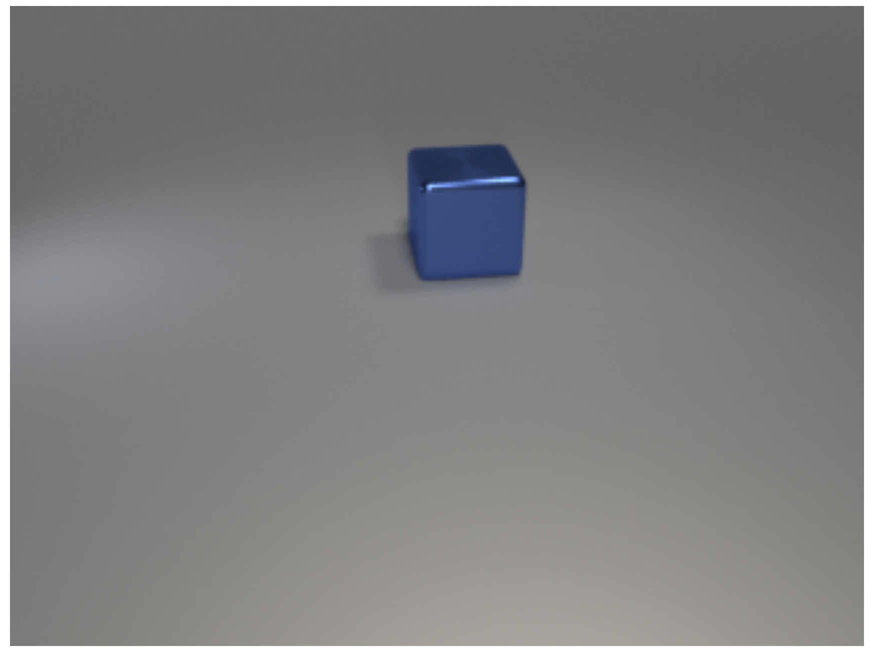
- Add, Mult, Conv
- Type-logical: $\underline{yellow}\ \overrightarrow{sphere}$
- Role-filler:

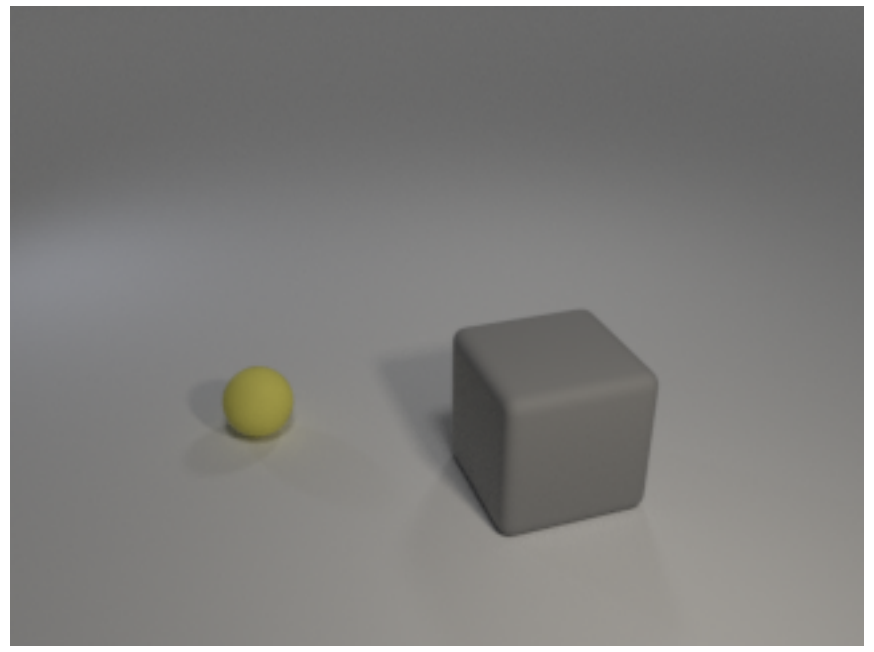
$\overrightarrow{yellow} \circledast \overrightarrow{adj} + \overrightarrow{sphere} \circledast \overrightarrow{noun}$



$$J(\theta) = -\log \sigma(\langle m, f(w_1, \ldots, w_n)\rangle) - \sum_{\bar{p} \in P} \log \sigma(-\langle m, f(w_{\bar{p}1}, \ldots, w_{\bar{p}n})\rangle)$$

Results - Single Object and Two Object

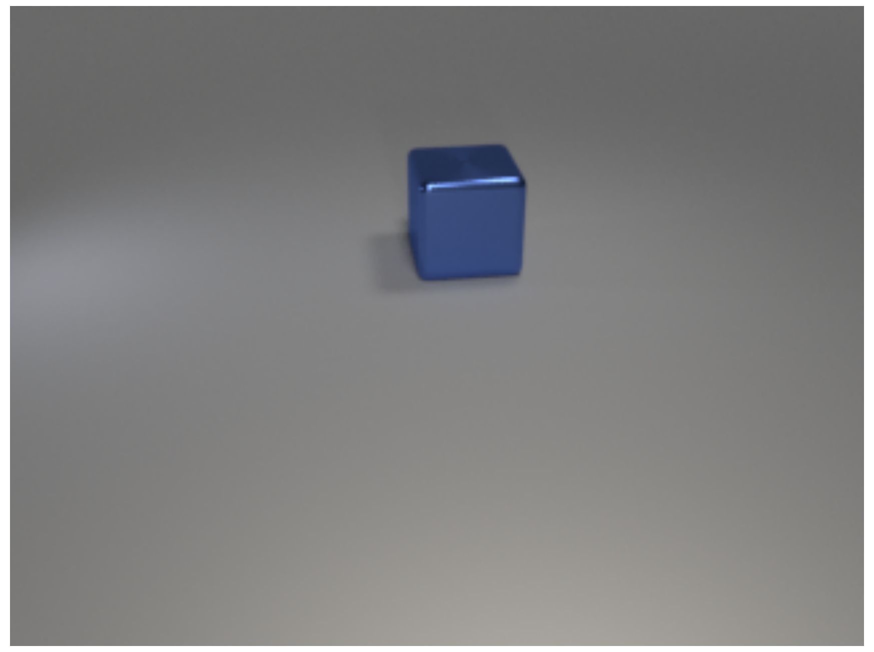


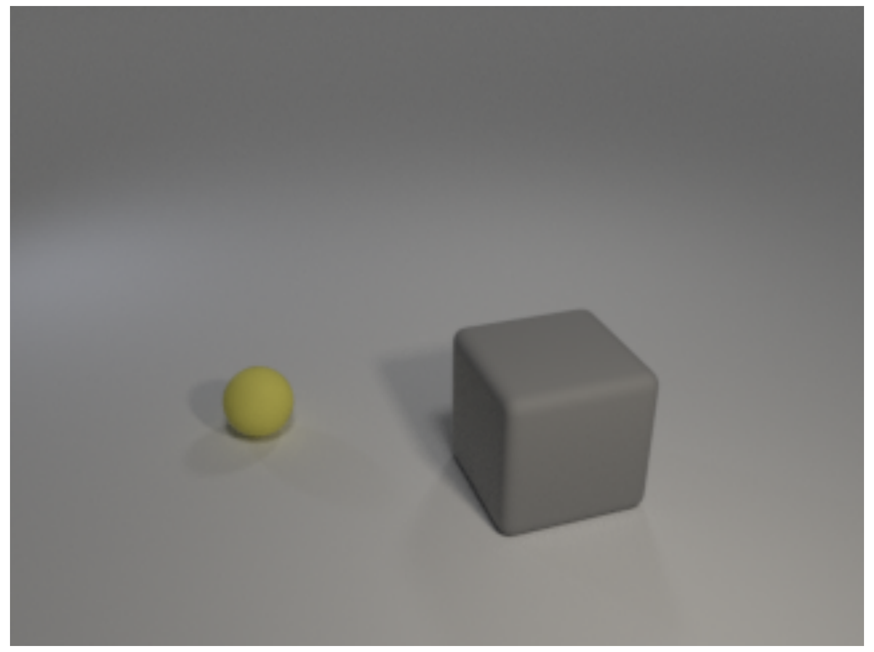


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	$98.98_{1.02}$	$89.06_{5.84}$	$78.54_{4.41}$
CSP	$94.98_{0.45}$	$84.58_{0.16}$	$88.74_{0.34}$
Add	$99.77_{0.03}$	$44.98_{1.32}$	$85.16_{0.96}$
Mult	$43.27_{13.9}$	$4.48_{4.08}$	$5.38_{2.66}$
Conv	$41.10_{14.3}$	$7.33_{2.90}$	$4.11_{1.53}$
TL	$99.98_{0.02}$	$1.08_{0.44}$	$0.92_{0.24}$
RF	$98.87_{0.11}$	$59.52_{6.12}$	$80.64_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	$86.91_{8.15}$	$6.31_{3.31}$	$0.25_{0.10}$
CSP	$37.59_{1.54}$	$20.98_{0.22}$	$11.15_{2.03}$
Add	$32.46_{0.11}$	$15.38_{0.89}$	$21.37_{0.60}$
Mult	$86.65_{8.93}$	$4.66_{1.35}$	$0.13_{0.03}$
Conv	$46.26_{0.53}$	$7.11_{2.18}$	$0.28_{0.14}$
TL	$99.41_{0.17}$	$21.23_{4.08}$	$0.08_{0.07}$
RF	$25.23_{1.08}$	$25.13_{3.99}$	$20.36_{1.36}$

Results - Single Object and Two Object

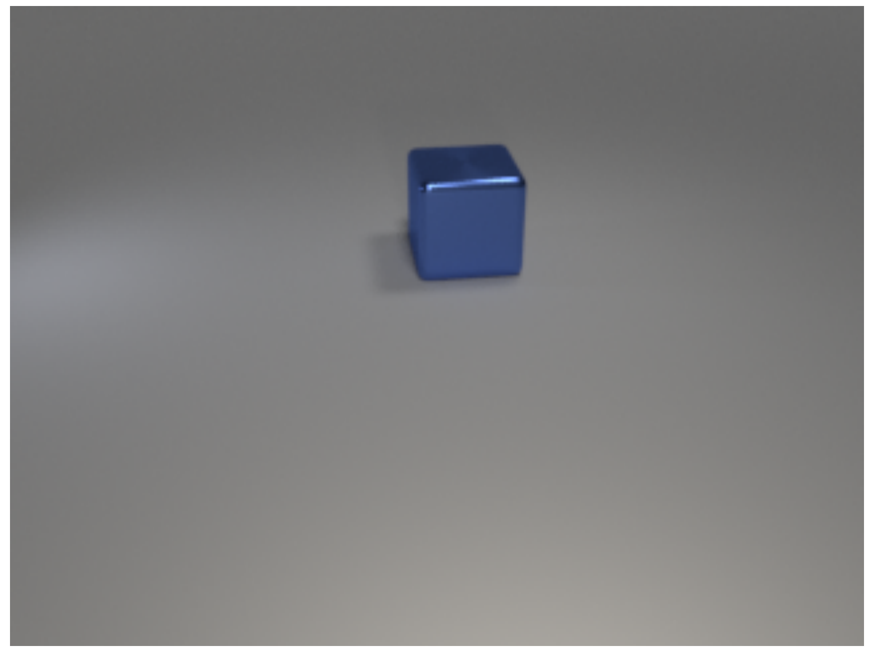


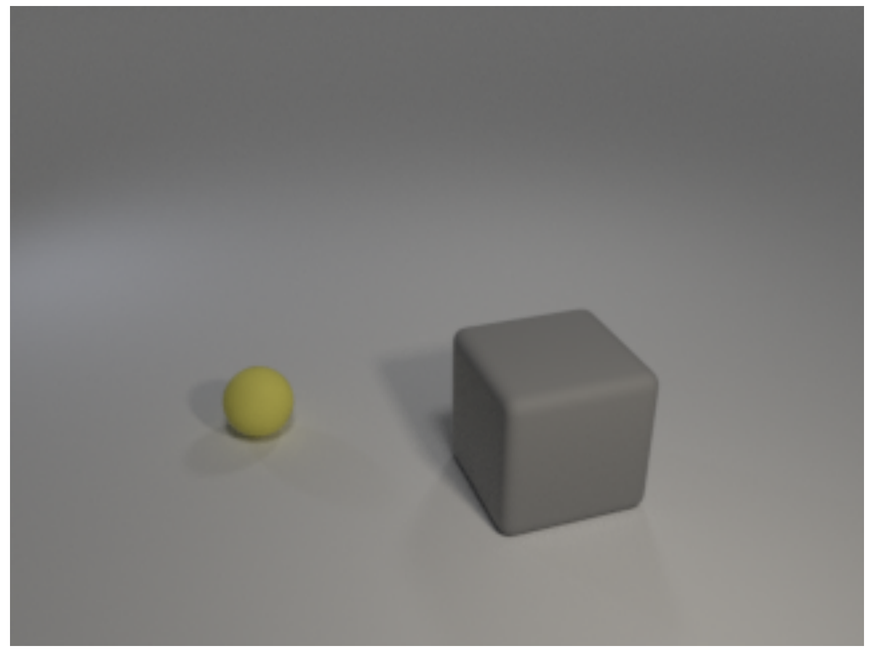


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	98.98 $_{1.02}$	89.06 $_{5.84}$	78.54 $_{4.41}$
CSP	94.98 $_{0.45}$	84.58 $_{0.16}$	88.74 $_{0.34}$
Add	99.77 $_{0.03}$	44.98 $_{1.32}$	85.16 $_{0.96}$
Mult	43.27 $_{13.9}$	4.48 $_{4.08}$	5.38 $_{2.66}$
Conv	41.10 $_{14.3}$	7.33 $_{2.90}$	4.11 $_{1.53}$
TL	99.98 $_{0.02}$	1.08 $_{0.44}$	0.92 $_{0.24}$
RF	98.87 $_{0.11}$	59.52 $_{6.12}$	80.64 $_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	86.91 $_{8.15}$	6.31 $_{3.31}$	0.25 $_{0.10}$
CSP	37.59 $_{1.54}$	20.98 $_{0.22}$	11.15 $_{2.03}$
Add	32.46 $_{0.11}$	15.38 $_{0.89}$	21.37 $_{0.60}$
Mult	86.65 $_{8.93}$	4.66 $_{1.35}$	0.13 $_{0.03}$
Conv	46.26 $_{0.53}$	7.11 $_{2.18}$	0.28 $_{0.14}$
TL	99.41 $_{0.17}$	21.23 $_{4.08}$	0.08 $_{0.07}$
RF	25.23 $_{1.08}$	25.13 $_{3.99}$	20.36 $_{1.36}$

Results - Single Object and Two Object

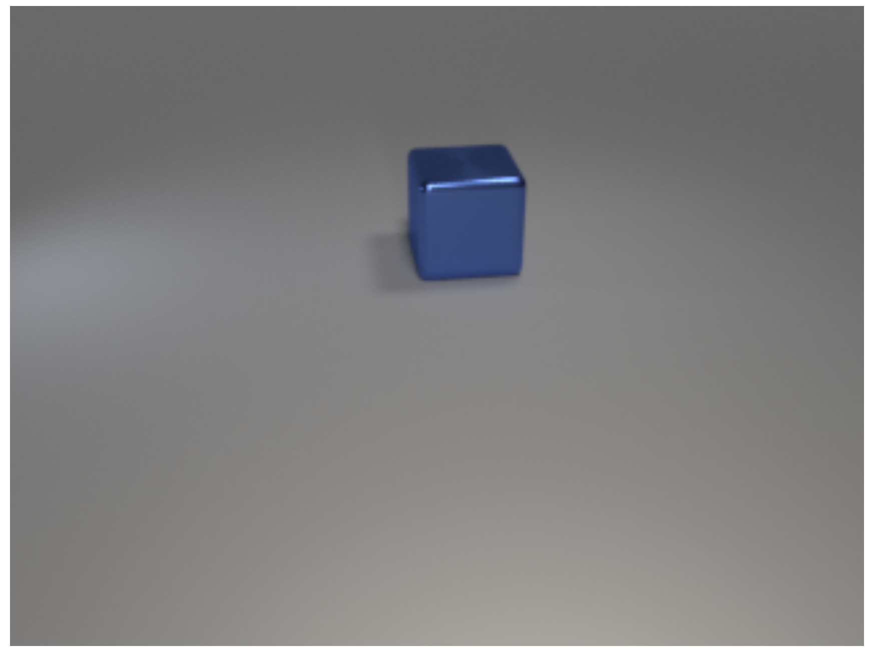


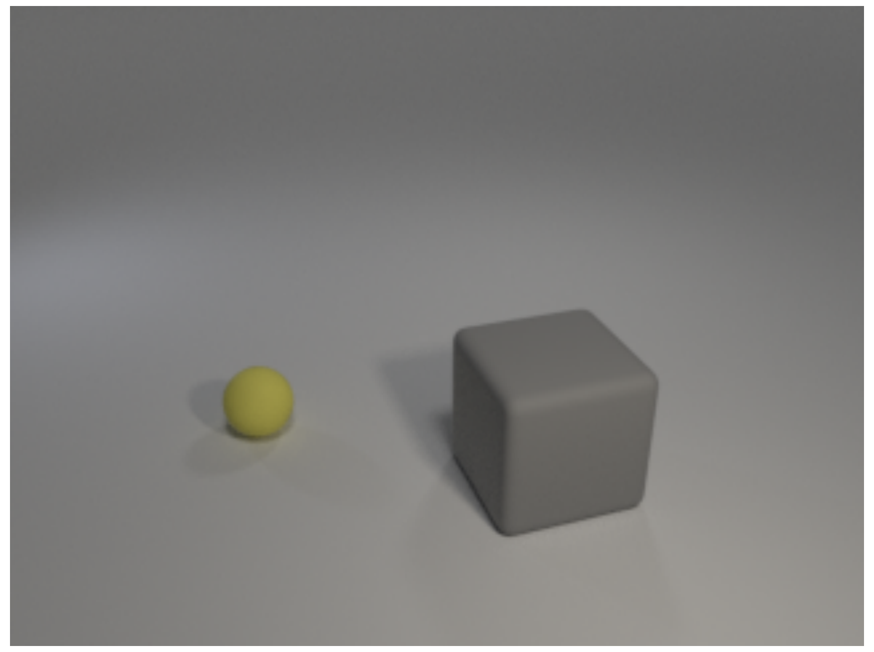


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	98.98 $_{1.02}$	89.06 $_{5.84}$	78.54 $_{4.41}$
CSP	94.98 $_{0.45}$	84.58 $_{0.16}$	88.74 $_{0.34}$
Add	99.77 $_{0.03}$	44.98 $_{1.32}$	85.16 $_{0.96}$
Mult	43.27 $_{13.9}$	4.48 $_{4.08}$	5.38 $_{2.66}$
Conv	41.10 $_{14.3}$	7.33 $_{2.90}$	4.11 $_{1.53}$
TL	99.98 $_{0.02}$	1.08 $_{0.44}$	0.92 $_{0.24}$
RF	98.87 $_{0.11}$	59.52 $_{6.12}$	80.64 $_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	86.91 $_{8.15}$	6.31 $_{3.31}$	0.25 $_{0.10}$
CSP	37.59 $_{1.54}$	20.98 $_{0.22}$	11.15 $_{2.03}$
Add	32.46 $_{0.11}$	15.38 $_{0.89}$	21.37 $_{0.60}$
Mult	86.65 $_{8.93}$	4.66 $_{1.35}$	0.13 $_{0.03}$
Conv	46.26 $_{0.53}$	7.11 $_{2.18}$	0.28 $_{0.14}$
TL	99.41 $_{0.17}$	21.23 $_{4.08}$	0.08 $_{0.07}$
RF	25.23 $_{1.08}$	25.13 $_{3.99}$	20.36 $_{1.36}$

Results - Single Object and Two Object

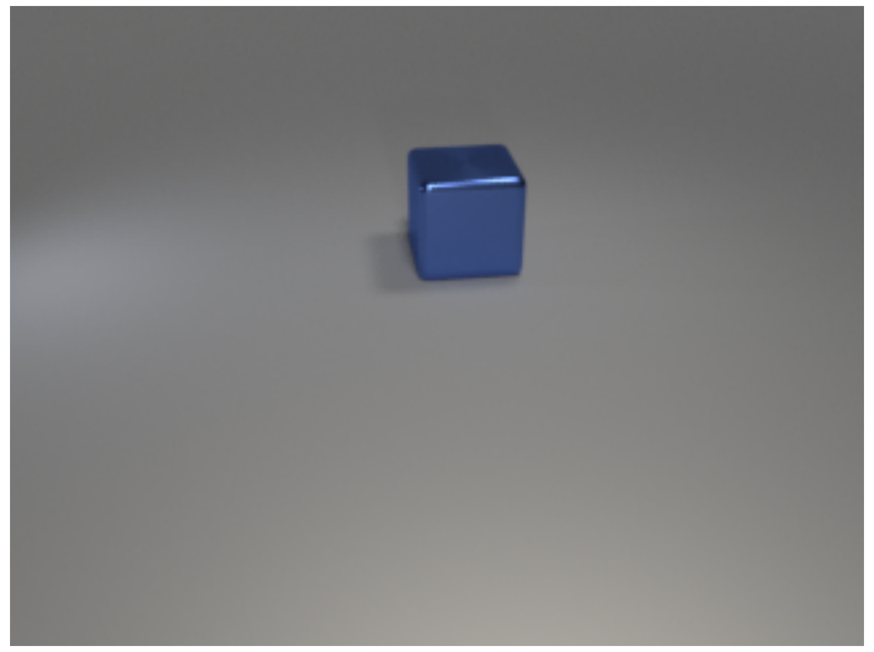


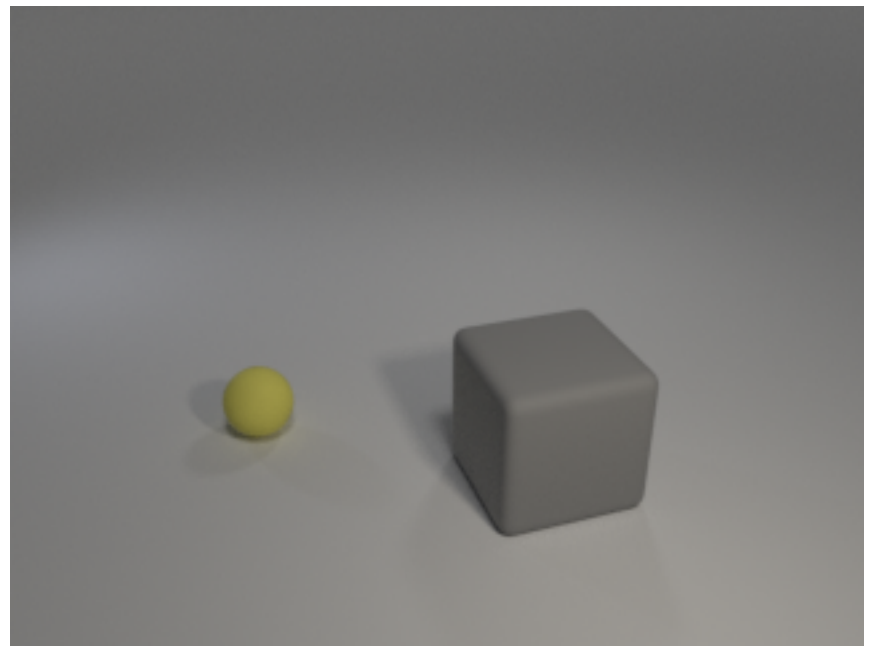


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	98.98 $_{1.02}$	89.06 $_{5.84}$	78.54 $_{4.41}$
CSP	94.98 $_{0.45}$	84.58 $_{0.16}$	88.74 $_{0.34}$
Add	99.77 $_{0.03}$	44.98 $_{1.32}$	85.16 $_{0.96}$
Mult	43.27 $_{13.9}$	4.48 $_{4.08}$	5.38 $_{2.66}$
Conv	41.10 $_{14.3}$	7.33 $_{2.90}$	4.11 $_{1.53}$
TL	99.98 $_{0.02}$	1.08 $_{0.44}$	0.92 $_{0.24}$
RF	98.87 $_{0.11}$	59.52 $_{6.12}$	80.64 $_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	86.91 $_{8.15}$	6.31 $_{3.31}$	0.25 $_{0.10}$
CSP	37.59 $_{1.54}$	20.98 $_{0.22}$	11.15 $_{2.03}$
Add	32.46 $_{0.11}$	15.38 $_{0.89}$	21.37 $_{0.60}$
Mult	86.65 $_{8.93}$	4.66 $_{1.35}$	0.13 $_{0.03}$
Conv	46.26 $_{0.53}$	7.11 $_{2.18}$	0.28 $_{0.14}$
TL	99.41 $_{0.17}$	21.23 $_{4.08}$	0.08 $_{0.07}$
RF	25.23 $_{1.08}$	25.13 $_{3.99}$	20.36 $_{1.36}$

Results - Single Object and Two Object

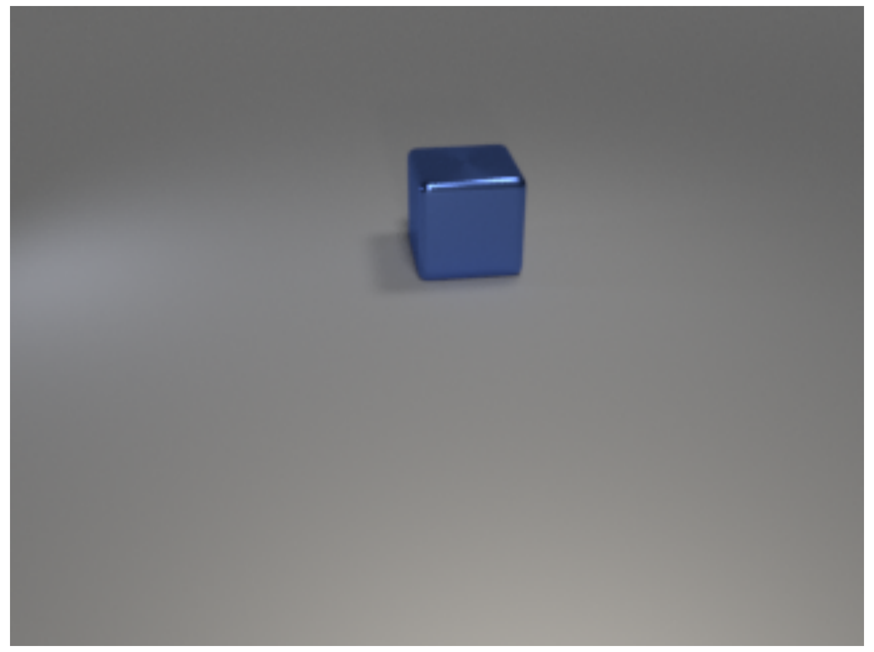


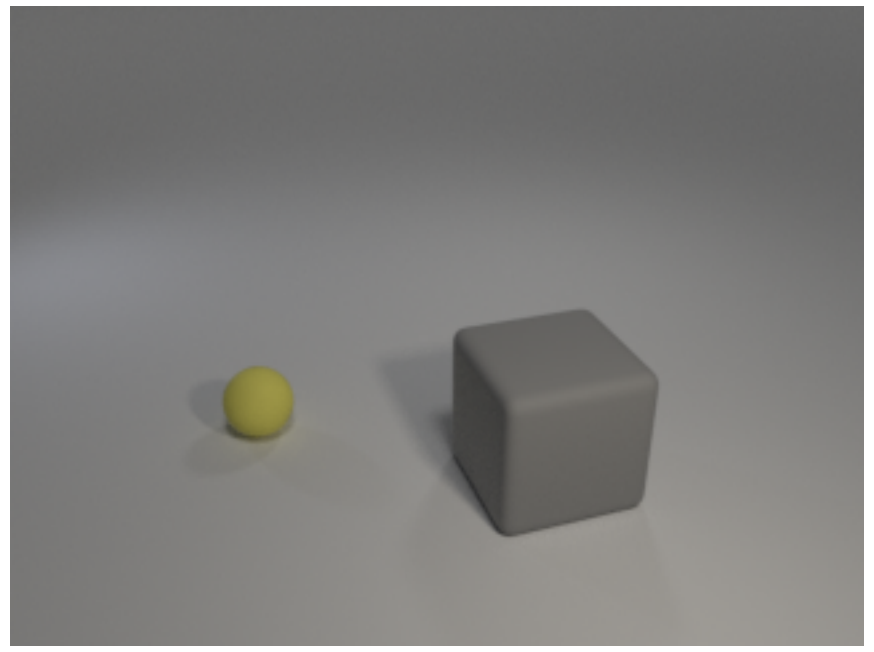


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	98.98 $_{1.02}$	89.06 $_{5.84}$	78.54 $_{4.41}$
CSP	94.98 $_{0.45}$	84.58 $_{0.16}$	88.74 $_{0.34}$
Add	99.77 $_{0.03}$	44.98 $_{1.32}$	85.16 $_{0.96}$
Mult	43.27 $_{13.9}$	4.48 $_{4.08}$	5.38 $_{2.66}$
Conv	41.10 $_{14.3}$	7.33 $_{2.90}$	4.11 $_{1.53}$
TL	99.98 $_{0.02}$	1.08 $_{0.44}$	0.92 $_{0.24}$
RF	98.87 $_{0.11}$	59.52 $_{6.12}$	80.64 $_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	86.91 $_{8.15}$	6.31 $_{3.31}$	0.25 $_{0.10}$
CSP	37.59 $_{1.54}$	20.98 $_{0.22}$	11.15 $_{2.03}$
Add	32.46 $_{0.11}$	15.38 $_{0.89}$	21.37 $_{0.60}$
Mult	86.65 $_{8.93}$	4.66 $_{1.35}$	0.13 $_{0.03}$
Conv	46.26 $_{0.53}$	7.11 $_{2.18}$	0.28 $_{0.14}$
TL	99.41 $_{0.17}$	21.23 $_{4.08}$	0.08 $_{0.07}$
RF	25.23 $_{1.08}$	25.13 $_{3.99}$	20.36 $_{1.36}$

Results - Single Object and Two Object

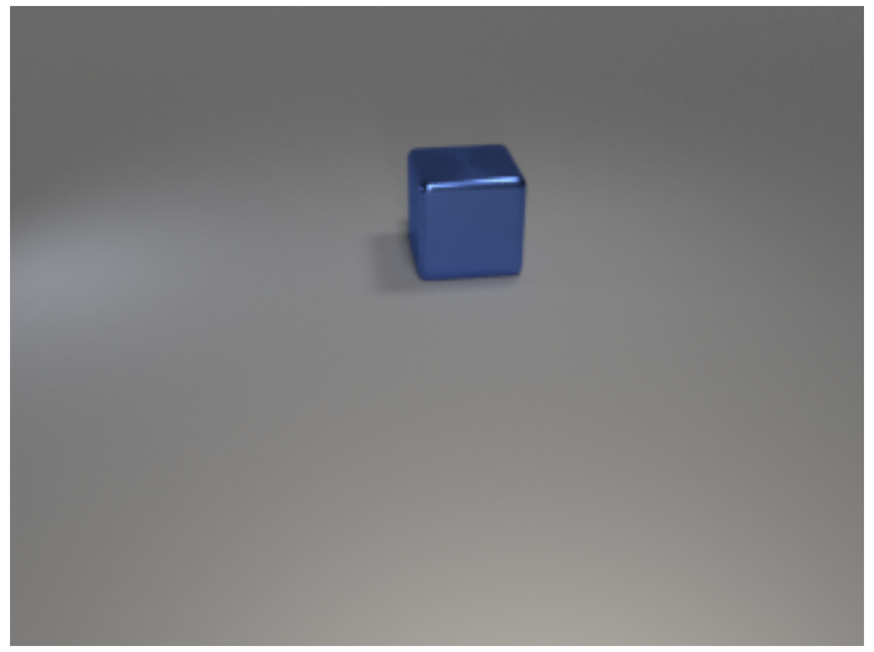


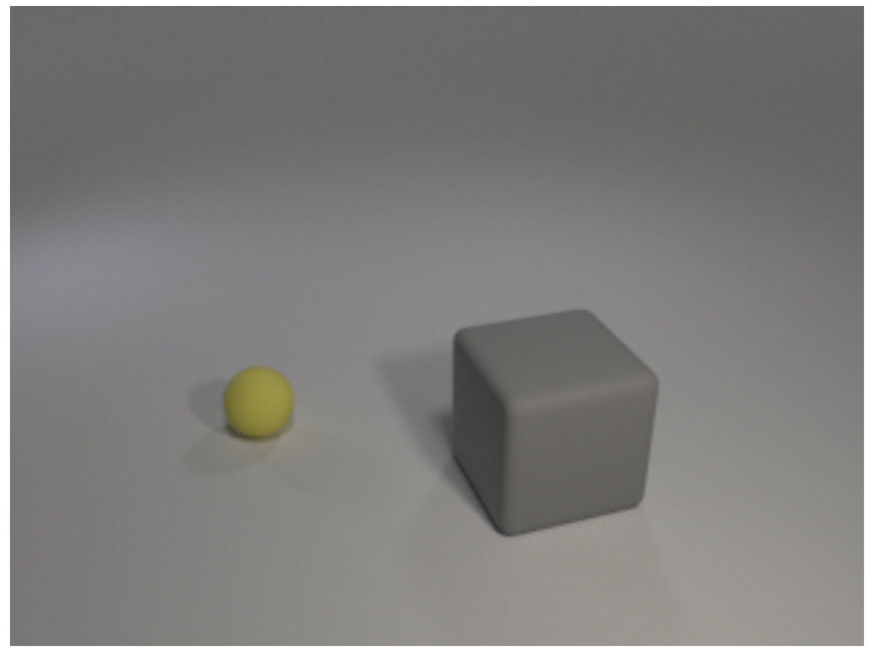


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	98.98 $_{1.02}$	89.06 $_{5.84}$	78.54 $_{4.41}$
CSP	94.98 $_{0.45}$	84.58 $_{0.16}$	88.74 $_{0.34}$
Add	99.77 $_{0.03}$	44.98 $_{1.32}$	85.16 $_{0.96}$
Mult	43.27 $_{13.9}$	4.48 $_{4.08}$	5.38 $_{2.66}$
Conv	41.10 $_{14.3}$	7.33 $_{2.90}$	4.11 $_{1.53}$
TL	99.98 $_{0.02}$	1.08 $_{0.44}$	0.92 $_{0.24}$
RF	98.87 $_{0.11}$	59.52 $_{6.12}$	80.64 $_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	86.91 $_{8.15}$	6.31 $_{3.31}$	0.25 $_{0.10}$
CSP	37.59 $_{1.54}$	20.98 $_{0.22}$	11.15 $_{2.03}$
Add	32.46 $_{0.11}$	15.38 $_{0.89}$	21.37 $_{0.60}$
Mult	86.65 $_{8.93}$	4.66 $_{1.35}$	0.13 $_{0.03}$
Conv	46.26 $_{0.53}$	7.11 $_{2.18}$	0.28 $_{0.14}$
TL	99.41 $_{0.17}$	21.23 $_{4.08}$	0.08 $_{0.07}$
RF	25.23 $_{1.08}$	25.13 $_{3.99}$	20.36 $_{1.36}$

Results - Single Object and Two Object

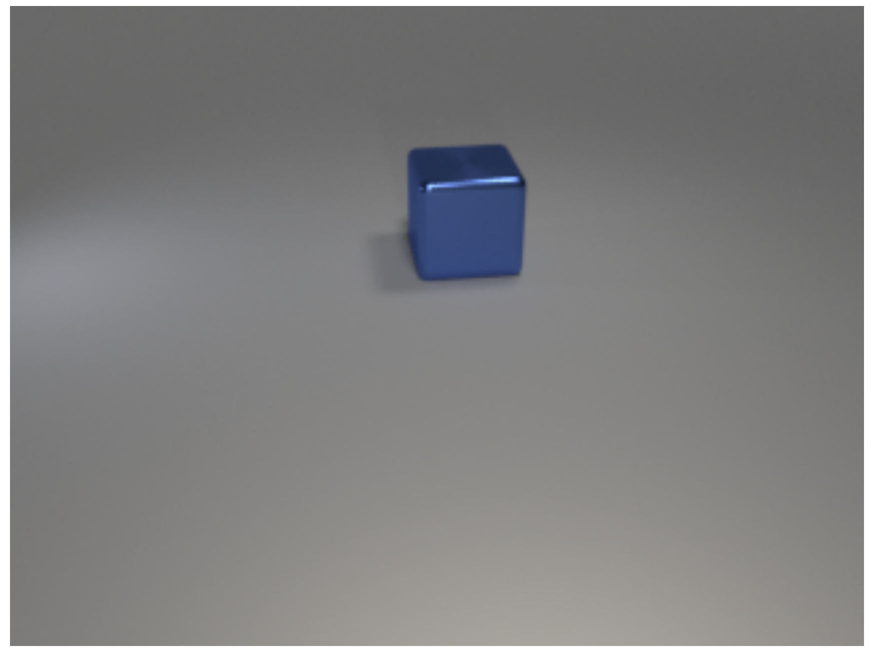


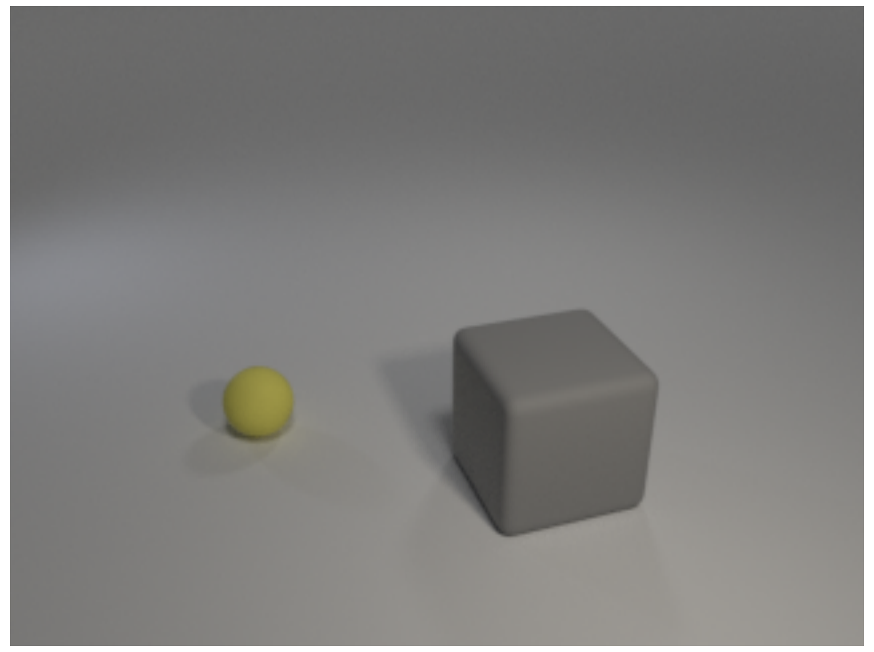


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	$98.98_{1.02}$	$89.06_{5.84}$	$78.54_{4.41}$
CSP	$94.98_{0.45}$	$84.58_{0.16}$	$88.74_{0.34}$
Add	$99.77_{0.03}$	$44.98_{1.32}$	$85.16_{0.96}$
Mult	$43.27_{13.9}$	$4.48_{4.08}$	$5.38_{2.66}$
Conv	$41.10_{14.3}$	$7.33_{2.90}$	$4.11_{1.53}$
TL	$99.98_{0.02}$	$1.08_{0.44}$	$0.92_{0.24}$
RF	$98.87_{0.11}$	$59.52_{6.12}$	$80.64_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	$86.91_{8.15}$	$6.31_{3.31}$	$0.25_{0.10}$
CSP	$37.59_{1.54}$	$20.98_{0.22}$	$11.15_{2.03}$
Add	$32.46_{0.11}$	$15.38_{0.89}$	$21.37_{0.60}$
Mult	$86.65_{8.93}$	$4.66_{1.35}$	$0.13_{0.03}$
Conv	$46.26_{0.53}$	$7.11_{2.18}$	$0.28_{0.14}$
TL	$99.41_{0.17}$	$21.23_{4.08}$	$0.08_{0.07}$
RF	$25.23_{1.08}$	$25.13_{3.99}$	$20.36_{1.36}$

Results - Single Object and Two Object

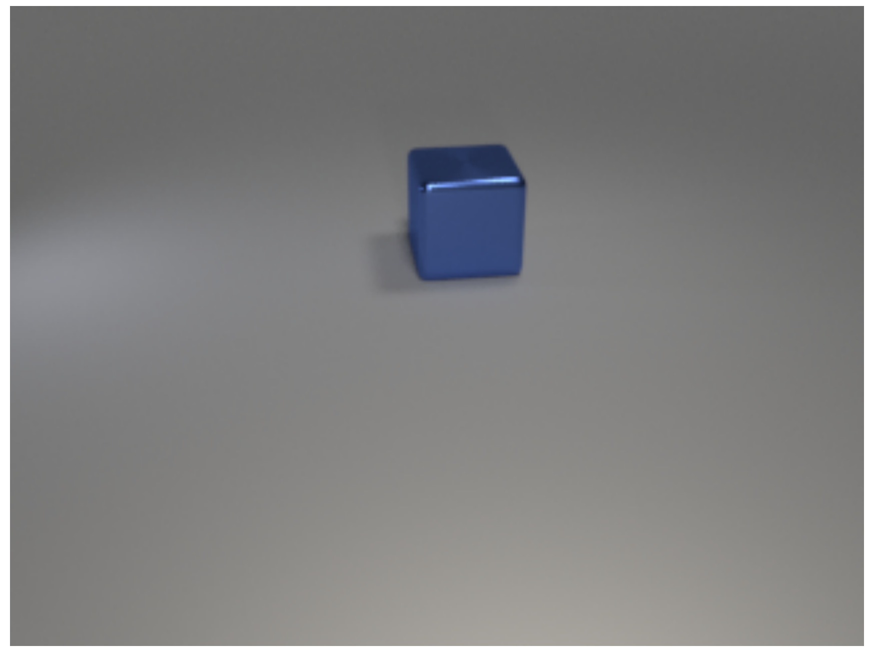




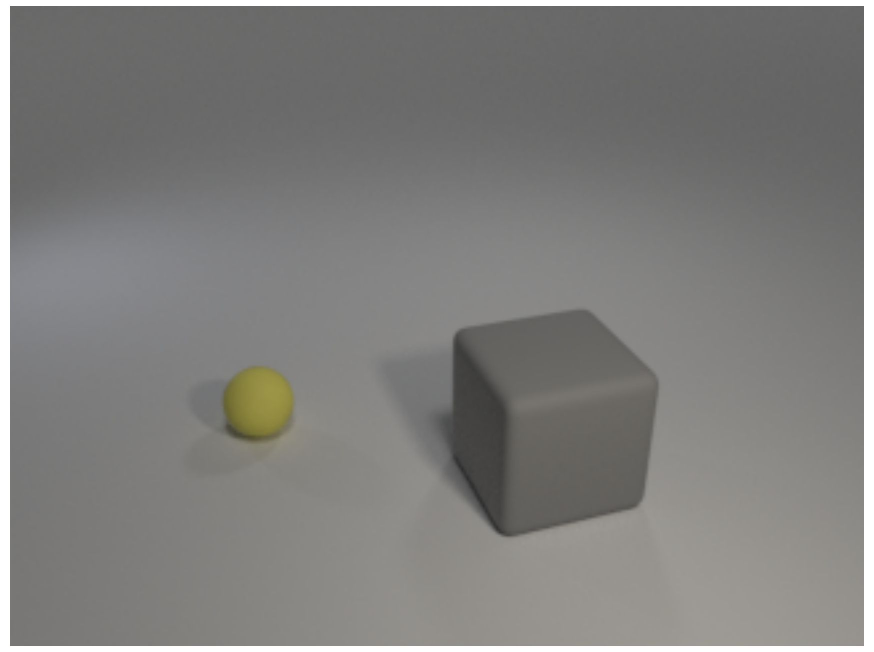
Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	98.98 $_{1.02}$	89.06 $_{5.84}$	78.54 $_{4.41}$
CSP	94.98 $_{0.45}$	84.58 $_{0.16}$	88.74 $_{0.34}$
Add	99.77 $_{0.03}$	44.98 $_{1.32}$	85.16 $_{0.96}$
Mult	43.27 $_{13.9}$	4.48 $_{4.08}$	5.38 $_{2.66}$
Conv	41.10 $_{14.3}$	7.33 $_{2.90}$	4.11 $_{1.53}$
TL	99.98 $_{0.02}$	1.08 $_{0.44}$	0.92 $_{0.24}$
RF	98.87 $_{0.11}$	59.52 $_{6.12}$	80.64 $_{1.36}$

Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	86.91 $_{8.15}$	6.31 $_{3.31}$	0.25 $_{0.10}$
CSP	37.59 $_{1.54}$	20.98 $_{0.22}$	11.15 $_{2.03}$
Add	32.46 $_{0.11}$	15.38 $_{0.89}$	21.37 $_{0.60}$
Mult	86.65 $_{8.93}$	4.66 $_{1.35}$	0.13 $_{0.03}$
Conv	46.26 $_{0.53}$	7.11 $_{2.18}$	0.28 $_{0.14}$
TL	99.41 $_{0.17}$	21.23 $_{4.08}$	0.08 $_{0.07}$
RF	25.23 $_{1.08}$	25.13 $_{3.99}$	20.36 $_{1.36}$

Results - Single Object and Two Object

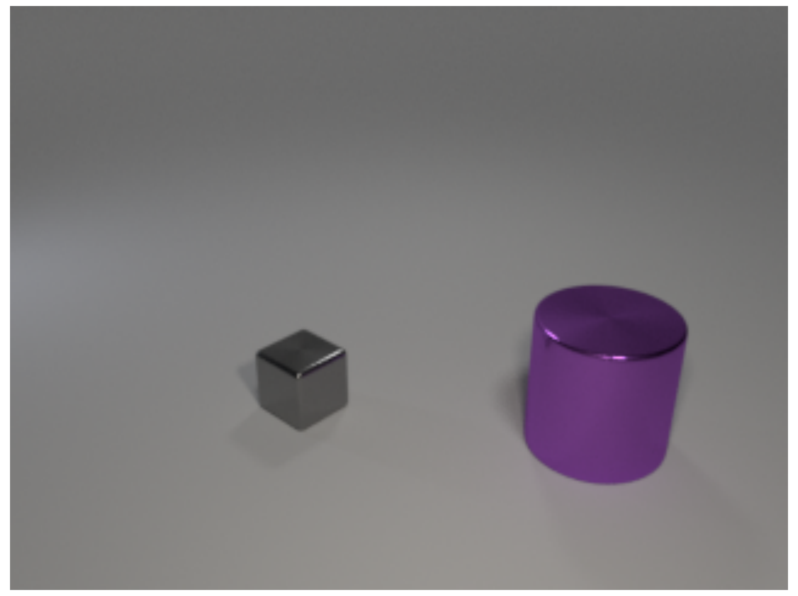


Model	Train	Val	Gen
CLIP	94.23	97.75	92.39
CLIP-FT	98.98 $_{1.02}$	89.06 $_{5.84}$	78.54 $_{4.41}$
CSP	94.98 $_{0.45}$	84.58 $_{0.16}$	88.74 $_{0.34}$
Add	99.77 $_{0.03}$	44.98 $_{1.32}$	85.16 $_{0.96}$
Mult	43.27 $_{13.9}$	4.48 $_{4.08}$	5.38 $_{2.66}$
Conv	41.10 $_{14.3}$	7.33 $_{2.90}$	4.11 $_{1.53}$
TL	99.98 $_{0.02}$	1.08 $_{0.44}$	0.92 $_{0.24}$
RF	98.87 $_{0.11}$	59.52 $_{6.12}$	80.64 $_{1.36}$



Model	Train	Val	Gen
CLIP	27.02	7.17	31.40
CLIP-FT	86.91 $_{8.15}$	6.31 $_{3.31}$	0.25 $_{0.10}$
CSP	37.59 $_{1.54}$	20.98 $_{0.22}$	11.15 $_{2.03}$
Add	32.46 $_{0.11}$	15.38 $_{0.89}$	21.37 $_{0.60}$
Mult	86.65 $_{8.93}$	4.66 $_{1.35}$	0.13 $_{0.03}$
Conv	46.26 $_{0.53}$	7.11 $_{2.18}$	0.28 $_{0.14}$
TL	99.41 $_{0.17}$	21.23 $_{4.08}$	0.08 $_{0.07}$
RF	25.23 $_{1.08}$	25.13 $_{3.99}$	20.36 $_{1.36}$

Results - Relational

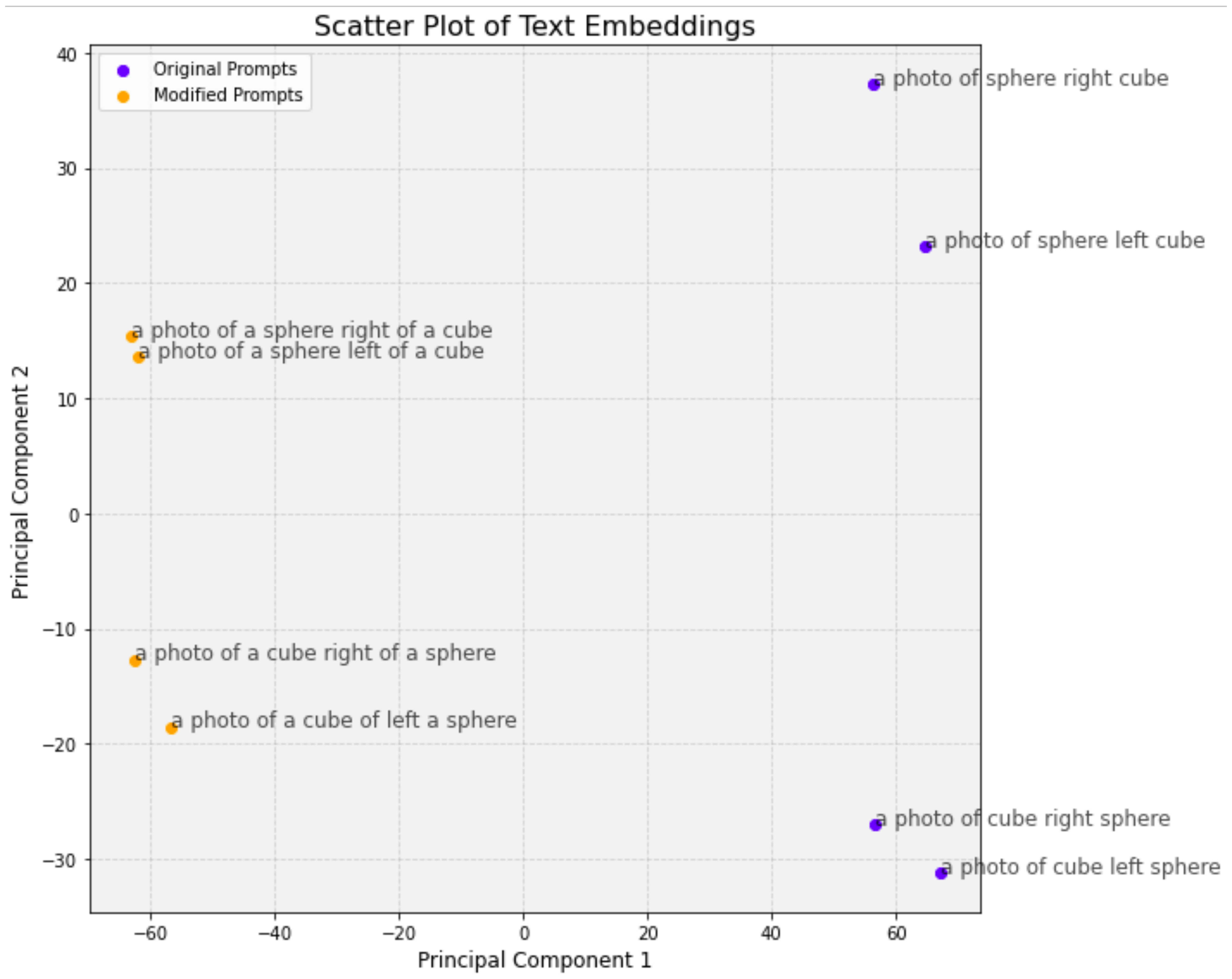


Model	Train	Val	Gen
CLIP	26.80	14.99	0.00
CLIP-FT	49.59 $_{0.44}$	0.00 $_{0.00}$	0.00 $_{0.00}$
CSP	30.40 $_{0.11}$	0.12 $_{0.01}$	0.03 $_{0.00}$
Add	25.41 $_{0.13}$	26.03 $_{0.07}$	25.47 $_{0.18}$
Mult	25.67 $_{0.12}$	25.95 $_{0.09}$	25.78 $_{0.09}$
Conv	24.83 $_{0.06}$	26.36 $_{0.55}$	24.95 $_{0.11}$
TL	99 ~~67.19~~ $_{0.26}$	0.00 $_{0.00}$	0.00 $_{0.00}$
RF	25.18 $_{0.28}$	24.89 $_{0.73}$	22.78 $_{0.20}$

- CLIP performs well in the single-object setting.
- Concept binding tasks hard for CLIP but also for compositional models.
- Relational task particularly hard
- Patterns of errors show differing performance even when overall accuracy is similar.

Lewis, Nayak et al., EACL Findings 2024,
https://aclanthology.org/2024.findings-eacl.101/

Text prompts are not adequately distinguished



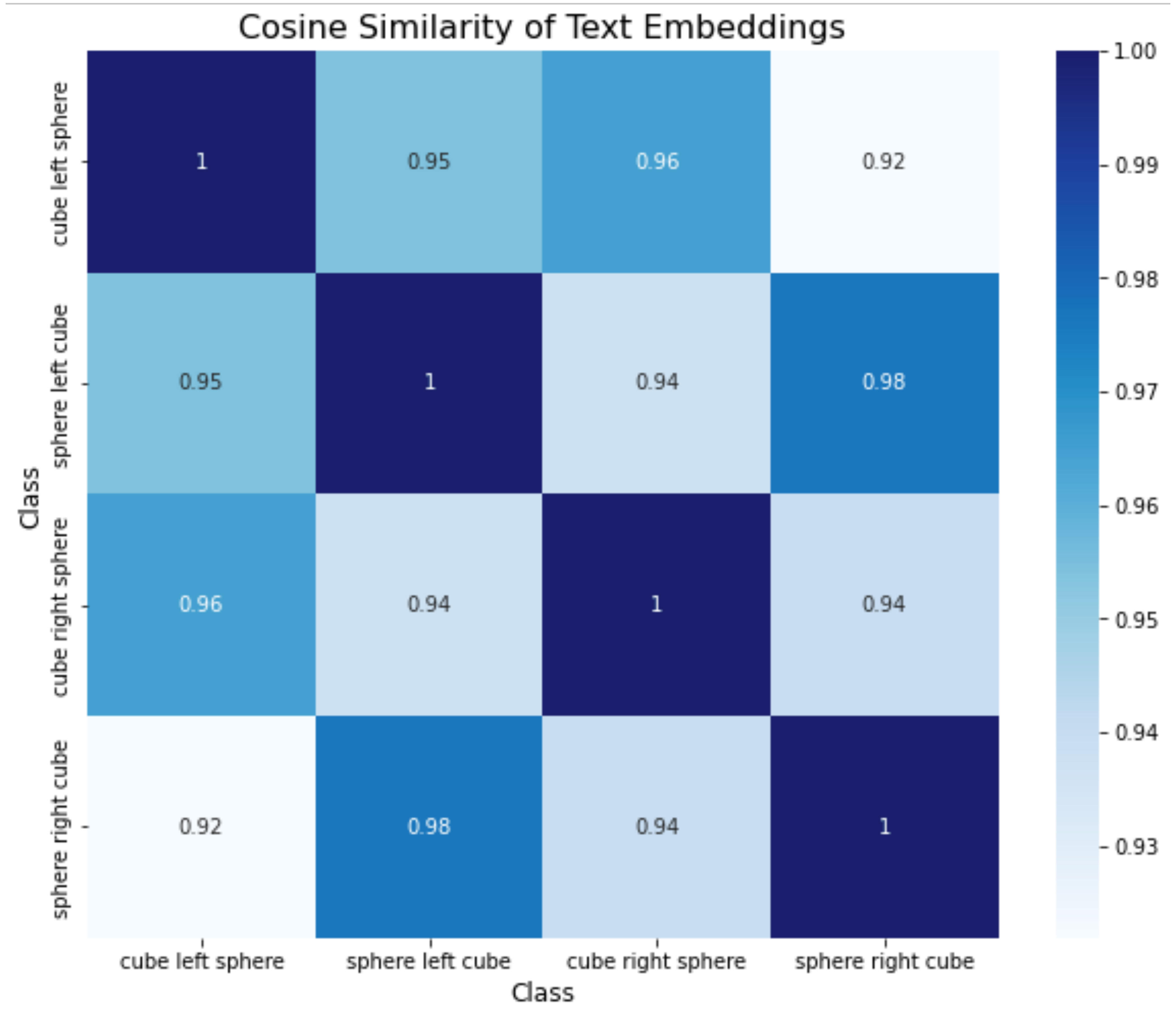
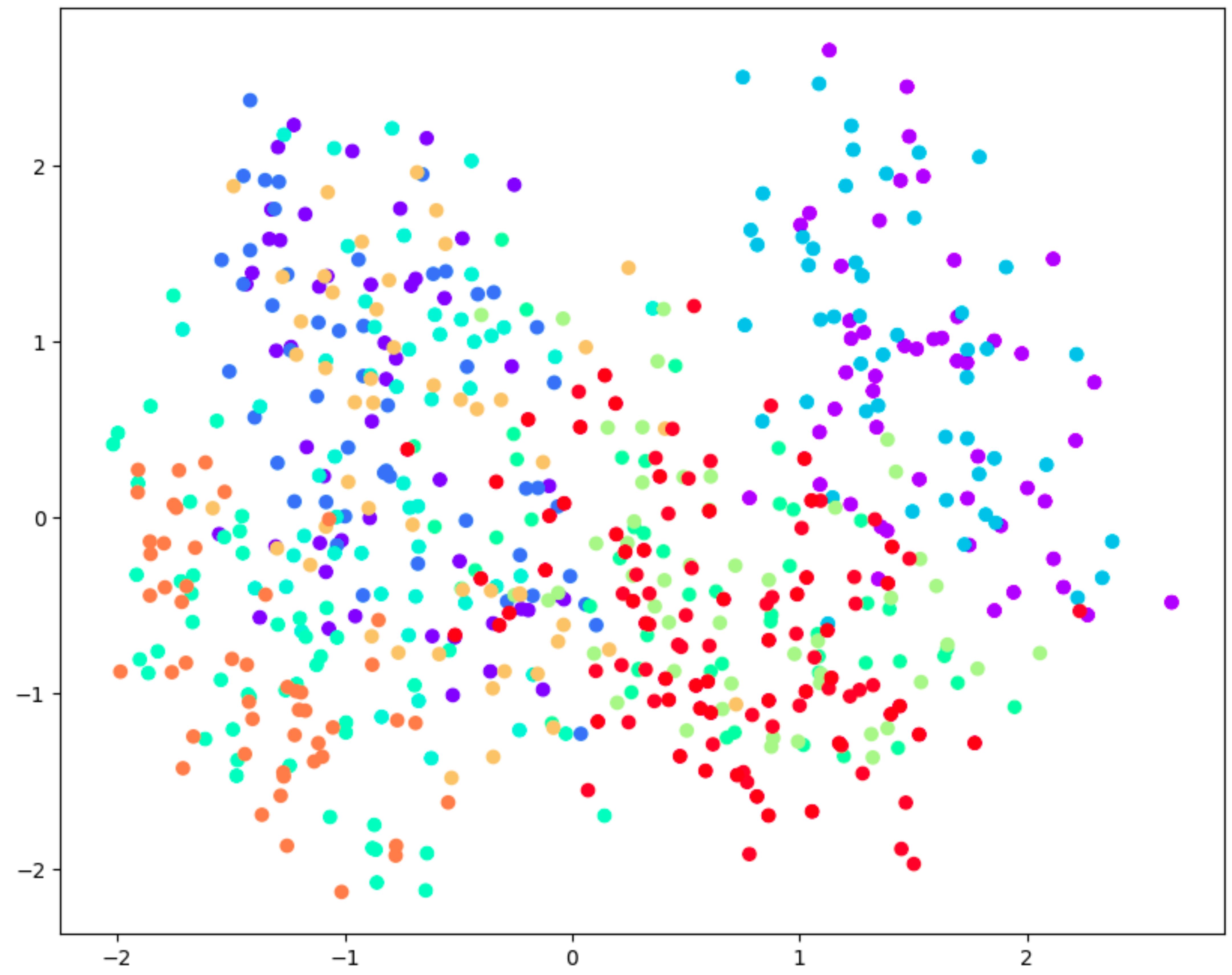
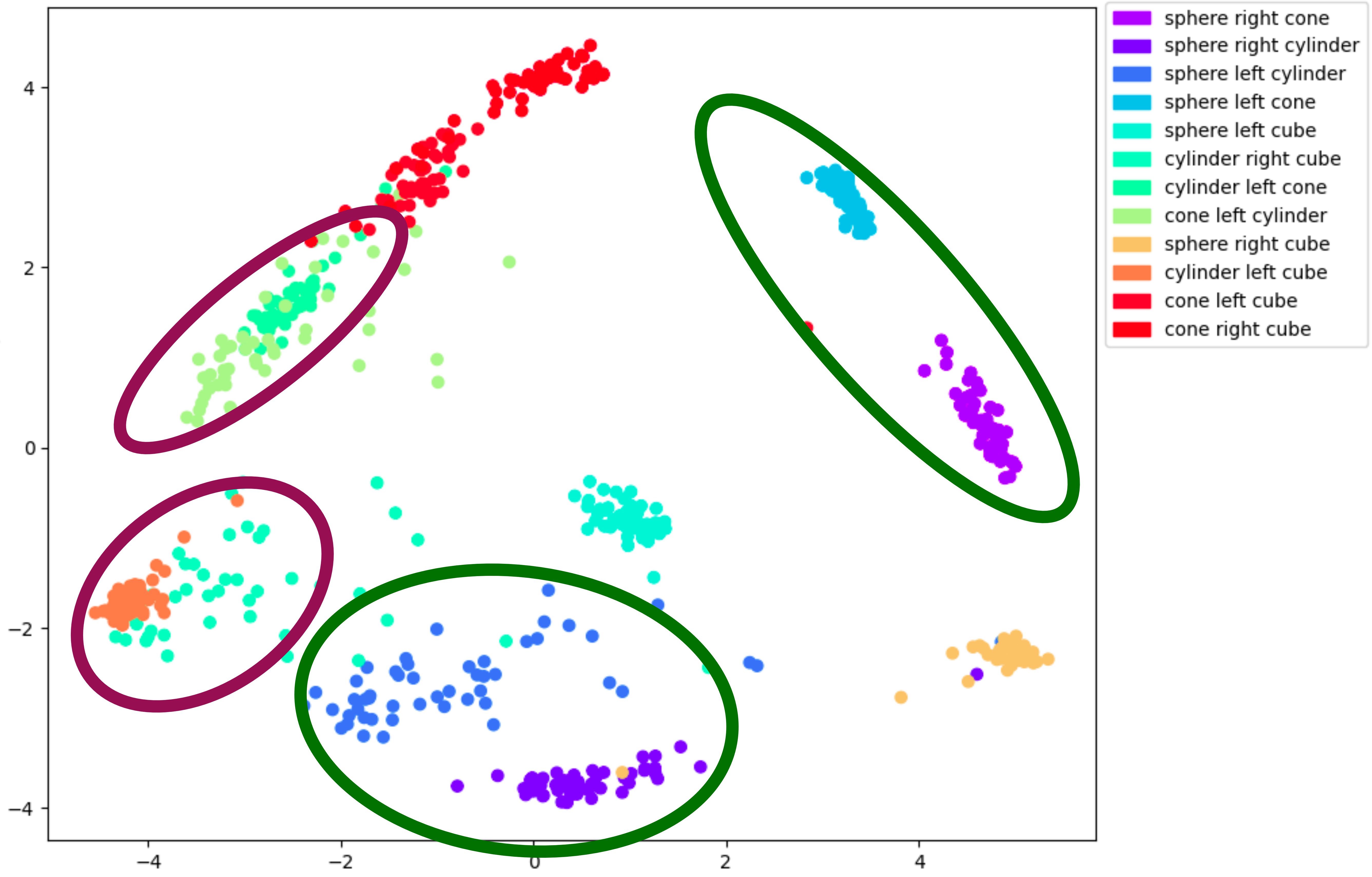


Image embeddings can also be ambiguous

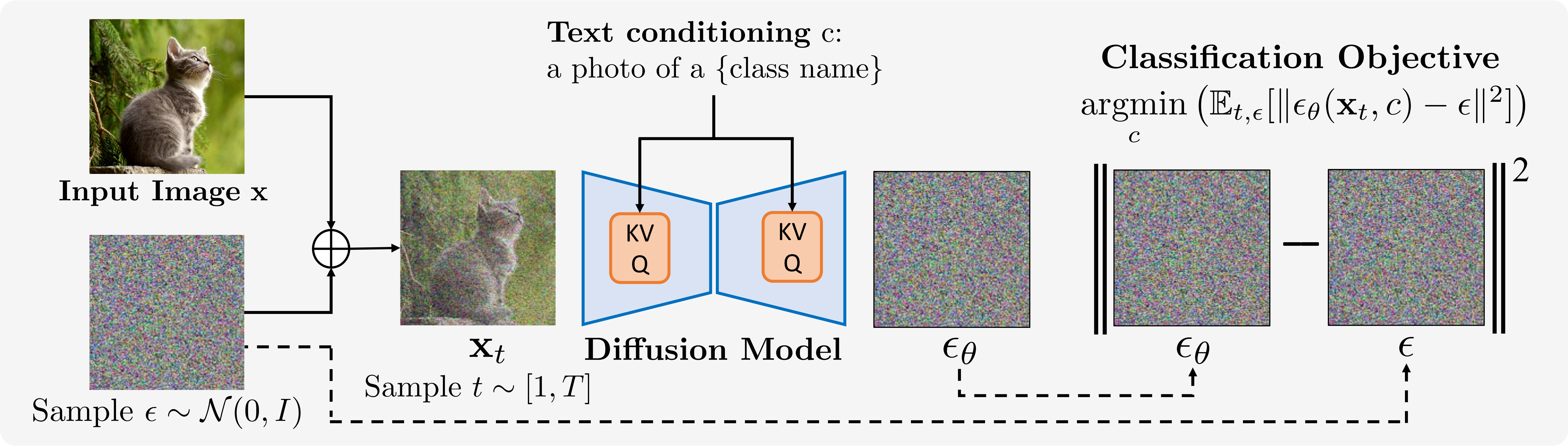
First two principal components - Frozen CLIP



First two principal components - Finetuned CLIP



Comparing with Diffusion Classifier



Model	Single Object	Two-Object ZS	Two-Object GZS
Frozen CLIP	99.5	93.0	35.3
CLIP-FT	100.0	99.7	17.2
Frozen DC	55.5	90.5	41.0
DC-FT	100.0	97.5	70.5

- Training only on single-object images.
- Between 30-50 training examples per class (vs. 1000's)
- What are the representations here?

Li et al, 2023, https://arxiv.org/abs/2303.16203
Pearson, Wray, Lewis, 2024

Summary and Further Work

- Developed novel methods to integrate type-logical and role-filler methods with deep neural architectures.
- CLIP is unable to generalize to unseen label combinations in a concept binding scenario. Compositional models do better on the training set.
- Diffusion Classifier seems promising in the GZSL setup.

Ambiguity and Metaphor

Words as density matrices

Meyer and Lewis CoNLL 2020 https://aclanthology.org/2020.conll-1.21/
Owers, Shutova, Lewis, QPL 2024 https://arxiv.org/abs/2408.11846v1

How should we deal with ambiguity?

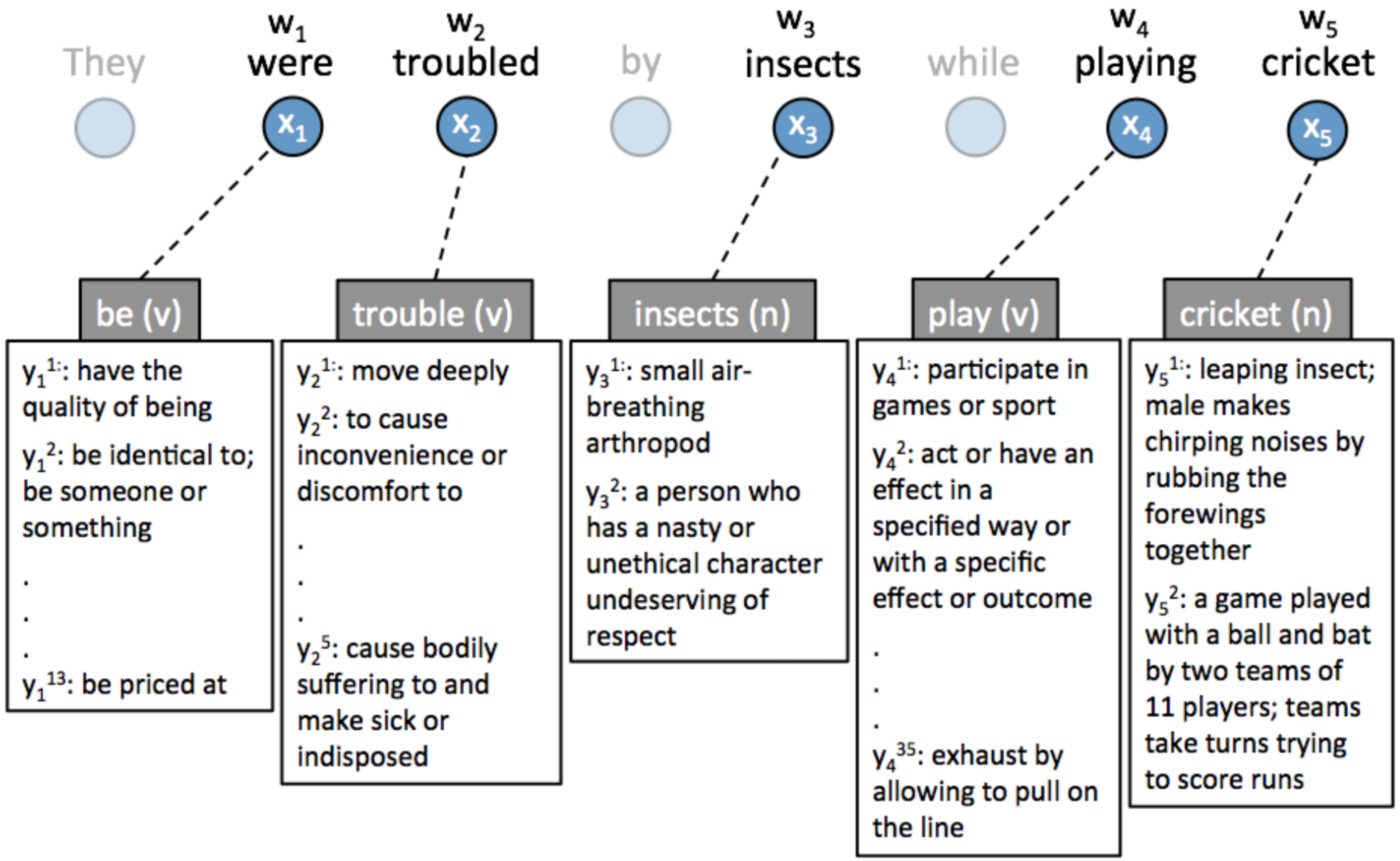


Figure 2: An example of the all-word WSD task. Content words and their possible senses are labeled w_i and y_i^j, respectively.

Chaplot, D. S., & Salakhutdinov, R. (2018). Knowledge-based word sense disambiguation using topic models. AAAI 18
Kartsaklis, D., & Sadrzadeh, M. (2013). "Prior disambiguation of word tensors for constructing sentence vectors." EMNLP 2013

Density matrices for word meaning

- A positive operator A on a (real) Hilbert space is a linear operator such that for any $|v\rangle$, the inner product $\langle v|A|v\rangle \geq 0$. A is self-adjoint and has positive eigenvalues.

- Given a word vector $|v\rangle$, we can lift it to the projection matrix $|v\rangle\langle v|$ associated with that vector.

$$|cat\rangle \mapsto |cat\rangle\langle cat|$$

- Given multiple senses of a word, we can combine their sense vectors together.

$$bed = p_r|bed_{river}\rangle\langle bed_{river}| + p_s|bed_{sleep}\rangle\langle bed_{sleep}|$$

- Words should disambiguate as they are **composed** in context.

Piedeleu et al., CALCO 2015

Neural density matrix embeddings

We learn multiple vectors for each word here

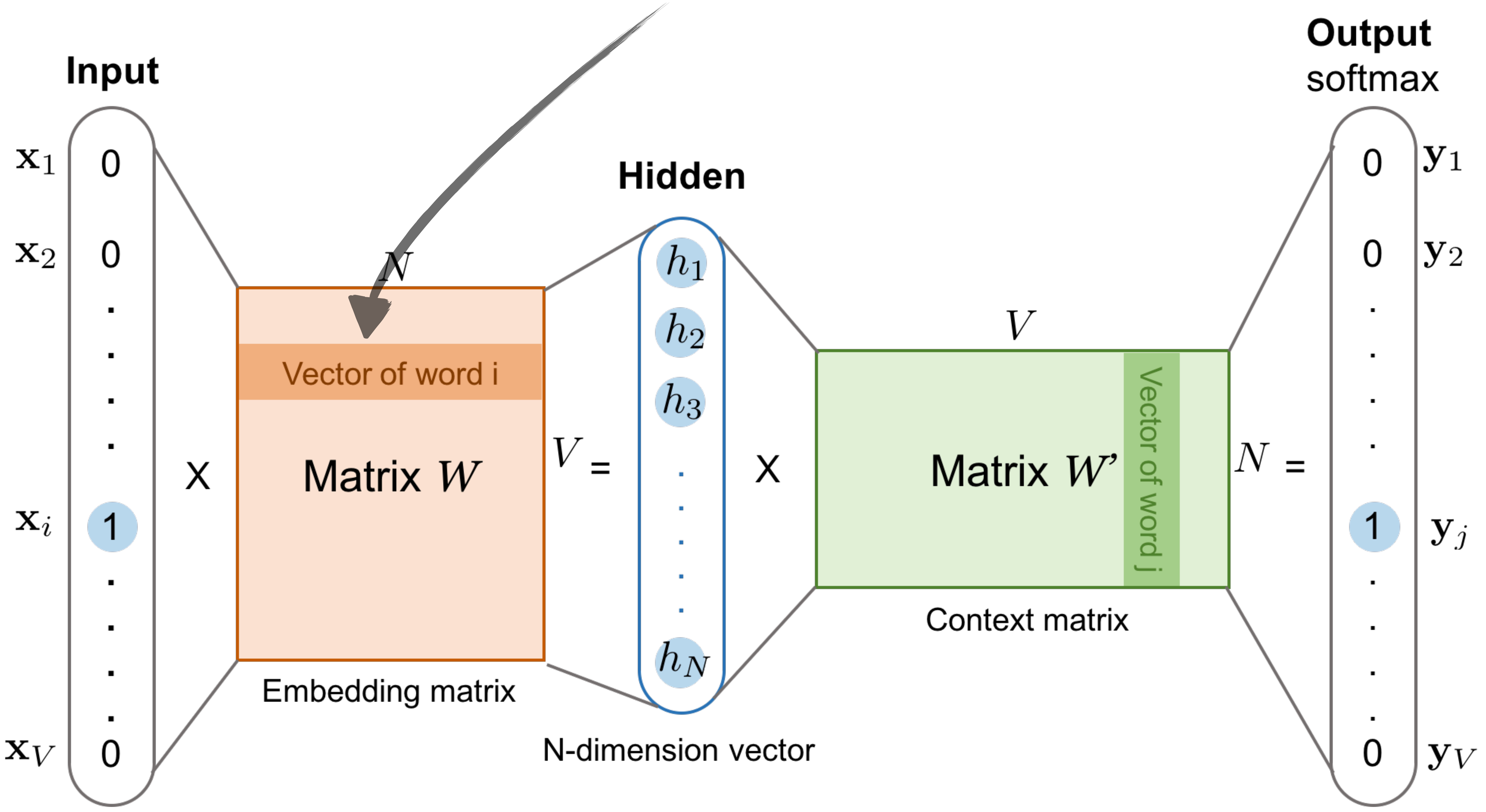


Image credit: https://lilianweng.github.io/posts/2017-10-15-word-embedding/

Task Description + Results

Density matrix methods beat state-of-the-art

Data set	Format	High similarity example	Low similarity example	Results	
ML2008	SV	value slump value decline	value slump value slouch	Glove Word2DM	0.397 0.328
GS2011	SVO	people buy house people purchase house	people buy house people bribe house	Glove Word2DM	0.304 **0.365**
GS2012	ASVAO	local family run small hotel local family operate small hotel	local family run small hotel local family move small hotel	BERT Word2DM	0.471 **0.500**
KS2013-CoNLL	ASVAO	young woman file long nail young woman smooth long nail	young woman file long nail young woman register long nail	BERT Word2DM	0.314 **0.345**

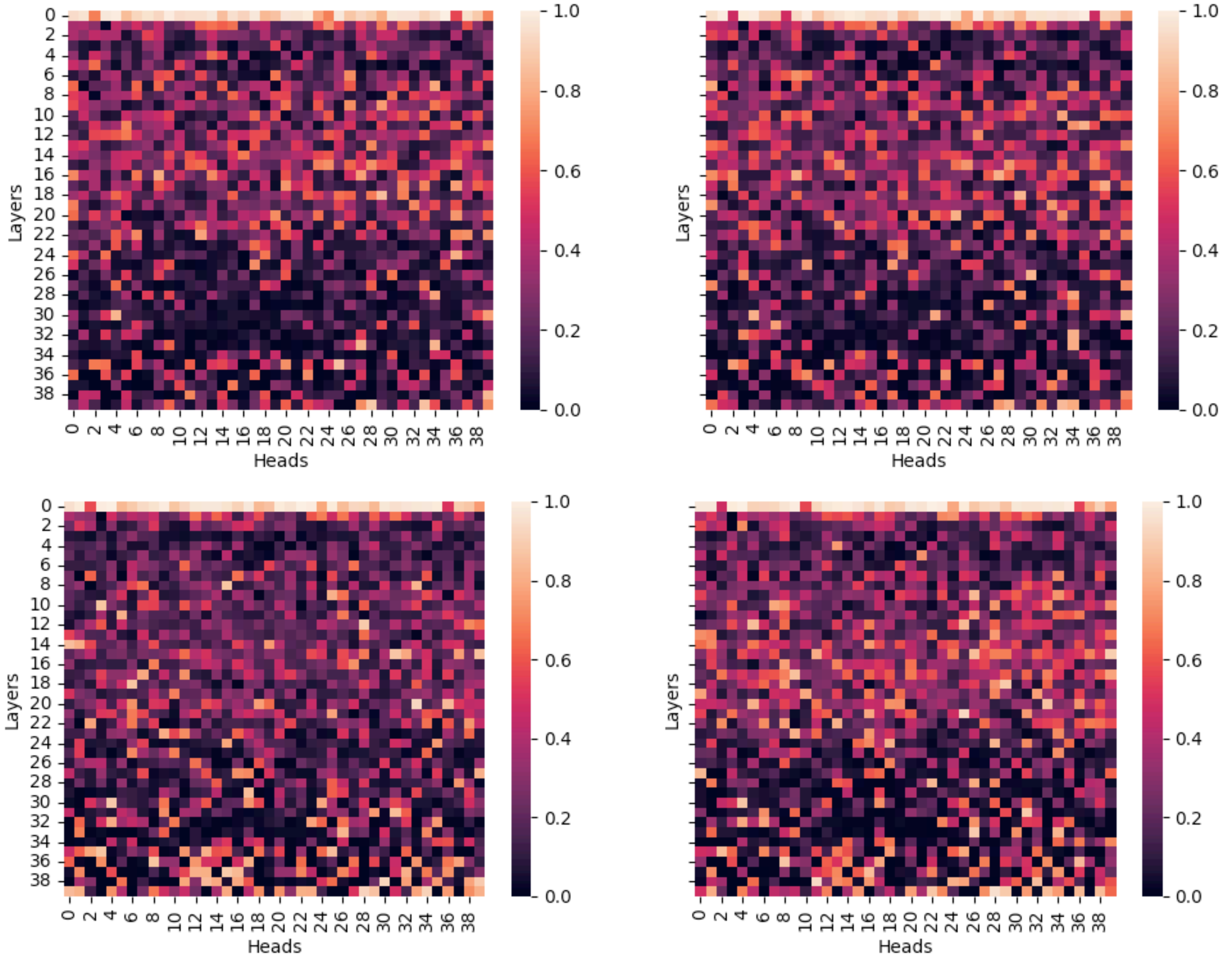
Summary

- We can use density matrices to model ambiguity.
- We can learn representations using neural methods, and plug these into a compositional/symbolic structure.
- Ambiguity resolves when density matrices are composed, although metaphor is more difficult.
- Future work: Quantum-inspired, already implemented in simulation — link with other QNLP techniques.

Future plans

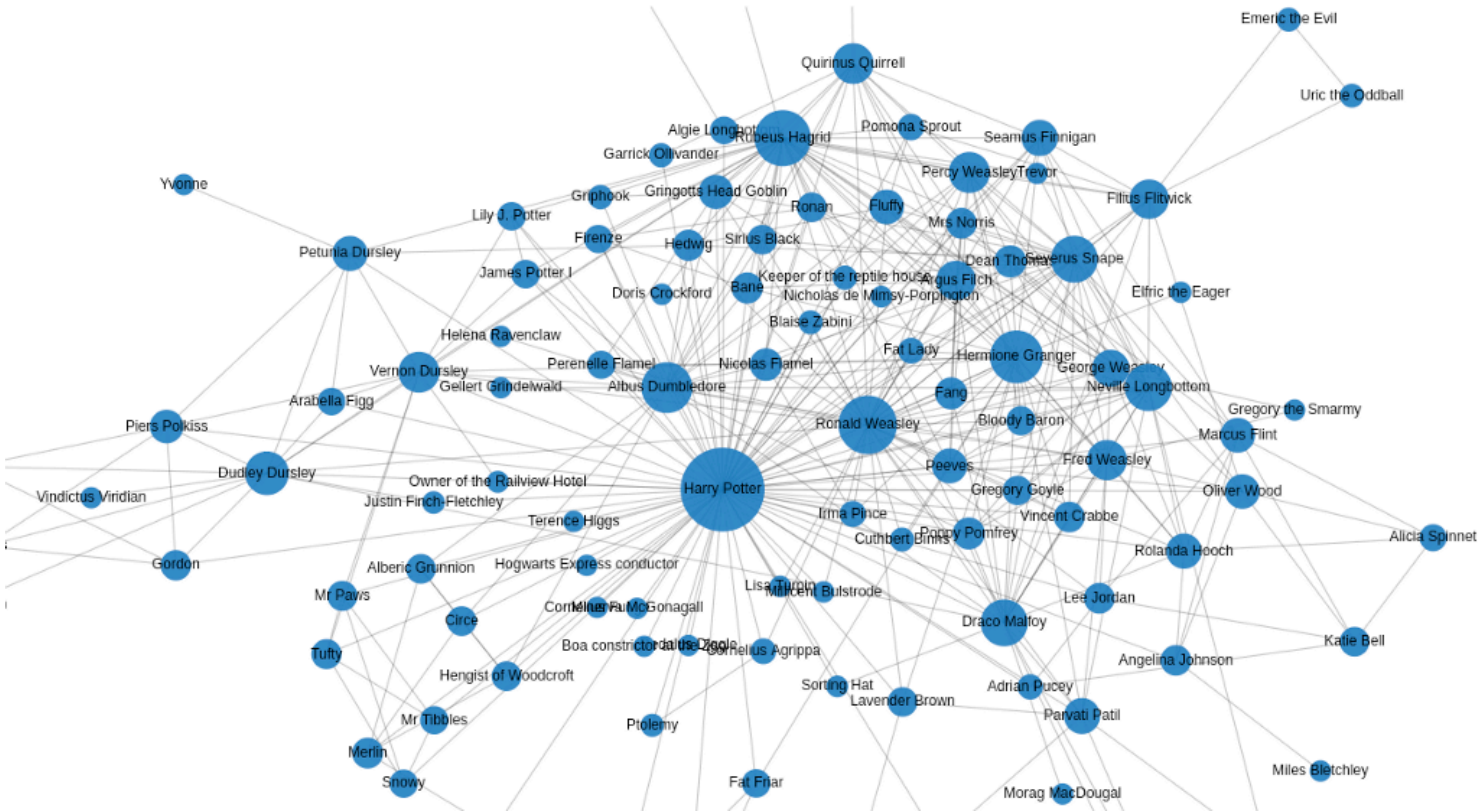
Understanding limitations of foundation models

… and how to improve them



- Understanding what kinds of representations are present.
- How representations might be used.
- How these could be leveraged to improve performance.

Exploring other forms of structured knowledge

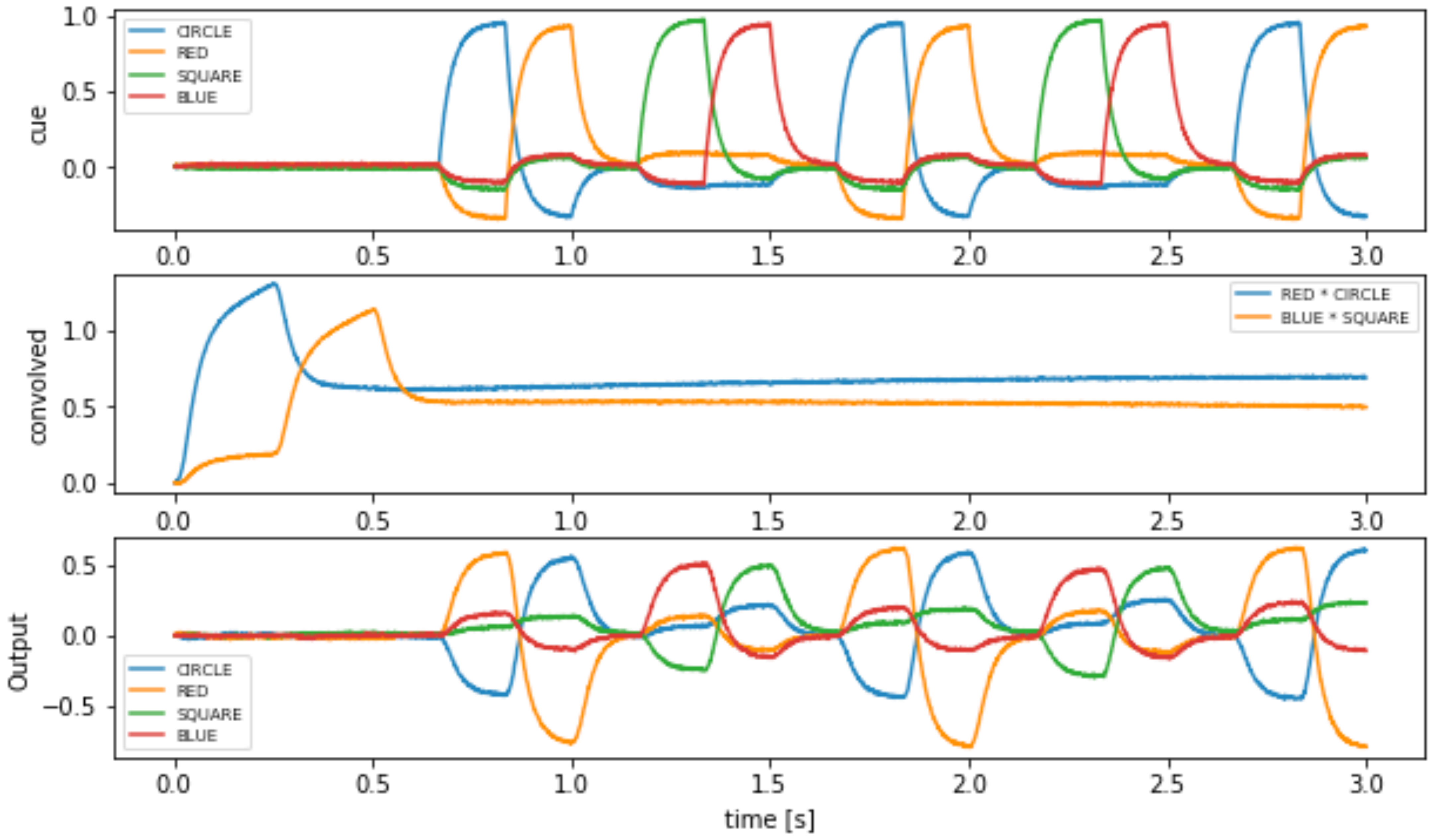


- RESCAL (Nickel et al., ICML 2011) has strong similarities to the type-logical model of meaning.
- WIP is exploring graph embedding techniques with Transformers.

Image credit: https://neo4j.com/developer-blog/turn-a-harry-potter-book-into-a-knowledge-graph/

Biologically 'realistic' neural networks

Nengo (Eliasmith, 2013)



- Drawbacks to role-filler models

- Mapping between role-filler and type-logical models (Lewis, https://arxiv.org/abs/2401.06808)

- Implemented a proof of concept in Nengo (video if time)

Summary

- Compositionality and flexibility are important aspects of human behaviour.
- At present, deep neural models are lacking.
- Compositional approaches may help in performance, and if not, may help in explainability and interpretability.
- We looked at analogical reasoning, visual reasoning, ambiguity.
- We looked at type-logical and role-filler models of composition.

Thank you for listening! I would love to hear questions or chat further!

marthaflinderslewis@gmail.com]

m.a.f.lewis@uva.nl

Thanks to students and collaborators!

Beth Pearson
Mike Wray
Nihal Nayak
Steve Bach
Ellie Pavlick
Melanie Mitchell
Katia Shutova
Jay Owers
Francois Meyer